An abstract approach to conditional independence in DIBI models

Anonymous Author(s)

Abstract

We propose two notions of conditional independence (CI) in a generic framework of copy-discard (CD) categories. Our abstract definitions instantiate to standard CI in probability theory and relational algebra, but also enable studying CI in non-standard settings such as multiset relations. They are formulated in the graphical formalism of string diagrams, which highlights the two notions' subtle difference when dealing with variables unrelated to the CI statement. Interestingly, while the two notions coincide in the probabilistic setting, the two notions differ in other CD categories, such as those formed by subprobability distributions and by non-normalised multisets.

As an application, we construct a class of categorical models for the logic of dependence and independence bunched implications (DIBI), and show that certain DIBI formula characterises exactly one of our new notions of CI. This class of categorical DIBI models is of independent interest, as it generalises the DIBI models in the original DIBI paper and facilitates building new ones.

CCS Concepts: • **Computer systems organization → Embedded systems**; *Redundancy*; Robotics; • **Networks →** Network reliability.

Keywords: DIBI logic, Categorical models, String diagrams, Conditional independence

ACM Reference Format:
Anonymous Author(s). 2018. An abstract approach to conditional independence in DIBI models. In *Proceedings of (Logic in Computer Science)*. ACM, New York, NY, USA, 18 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The notions of independence and conditional independence are useful in a variety of fields, including programming language theories[4], statistics [8], and database theory [2]. The general intuition is simple: two events A and B are *independent* if information about one does not provide any information about the other; they are *conditionally independent* (CI) given some other event C if given the information of C, there is no information flow between A and B.

However, reasoning about (conditional) independence is intricate [18], and there is a vast body of research studying how to formalise such reasoning. Among these works we identify two classes of techniques that have been put forward towards formalising CI: categorical and logical.

The categorical approaches define abstract notions of conditional independence. For instance, a series of diagrammatic definitions of CI have emerged since the pioneering work by Coecke and Spekkens [7]. Cho and Jacobs [5] presented CI as separation of 'boxes' in the graphical presentation of conditional probability distributions, which was later generalised by Fritz [10] to allow absence of conditionals in the categories. Jacobs and Zanasi [16] give a formal semantics in terms of Kleisli categories for d-separation in Bayesian reasoning. A different categorical setup deals instead with CI of generalised random variables [9, 11, 20]. The axioms in Simpson [20] do not define CI but serve as necessary conditions for any reasonable account of CI.

The logical approaches develop logical systems to express and axiomatise CI. Independence-Friendly Logic (IF) [12] and Dependence Logic [21] are two representatives that extend classical first-order logic with new quantifiers to describe (in)dependencies between variables. More recently, several program logics were designed to formally verify notions of (in)dependence in programs, especially in the context of security. For example, Barthe et al. [4] have introduced Probabilistic Separation Logic (PSL) to reason about the independence of variables generated by probabilistic programs, and applied their logic to verify the correctness of several cryptographic constructions. The assertion logic of PSL is the logic of Bunched Implications (BI) [17] but with a non-standard semantics in which models are Markov kernels and separation is interpreted as probabilistic independence.

Informally, one can think of the categorical approaches (in particular those via string diagrams) as 'semantic' as they describe the structure directly, and regard the logical approaches as 'syntactical' since they provide deduction systems for reasoning about CI.

Our paper brings together these two perspectives by providing categorical models for a logic describing dependency. We define two notions for CI on the same class of models: a categorical one (in terms of string diagrams), and a logical one (in terms of logic formulas). Identification of these two

Logic in Computer Science, August 02–05, 2022, Haifa, Israel

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

notions then provides us with two different ways to reason about the same phenomenon of CI.

In particular, we consider the Dependence and Independence Bunched Implication logic (DIBI), the assertion logic of a recently-proposed separation logic for probabilistic programs. Bao et al. [3] introduced Conditional Probabilistic Separation Logic (CPSL) as an extension of PSL to reason about conditional independence in probabilistic programs. This is achieved by extending the assertion logic from BI to DIBI by adding a non-commutative conjunction $\fatsemi$ and its adjoints (see Figure 1), thus enabling one to express conditional independence as independence plus sequential composition.

One reason for choosing DIBI is the generality of its models. In addition to the probabilistic models built up with Markov kernels, Bao et al. [3] also define a relational semantics for DIBI to reason about join-dependency — a notion of conditional independence in database theory and relational algebra. The probabilistic and relational models of DIBI share some properties and definitions in common. In particular, the states in both models are morphisms in some Kleisli categories (of the distribution monad and the non-empty powerset monad, respectively) that 'preserve inputs to outputs'. Such resemblance suggested the existence of a general class of models in which certain notions of conditional independence can be uniformly described via DIBI. Having such a general class of models would also overcome a big drawback of the setup in [3], where all the proofs had to be done independently for each model.

In this paper, we want to bridge the categorical and logical views of conditional independence and we achieve that through the following contributions:

1. We define a general class of DIBI models via categorical constructions.
   As Bao et al. [3] explain, the main difficulty is to define an essentially partial operation of parallel composition $\oplus$, such that the parallel composition of two morphisms (as states in the model) whose (co)domains may 'overlap' (e.g. $A \otimes B$ and $B \otimes C$) is a morphism whose (co)domain does not have duplicates (e.g. $A \otimes B \otimes C$ instead of $A \otimes B \otimes B \otimes C$). Our solution is to apply the string diagrammatic presentation of morphisms in symmetric monoidal categories (SMC), and eliminate the 'overlap' using certain copying structure.
2. As an application, we show how this categorical construction instantiates to the probabilistic and relational models in Bao et al. [3], and we provide a new multiset model for DIBI.
3. On the categorical side, we propose two notions of conditional independence in terms of string diagrams, reminiscent of that in Markov categories [10] yet tailored to our categorical setting.
4. We provide a detailed comparison to Simpson's categorical account of (conditional) independence as a (local) independence structure over abstract random variables [20] and show how a string diagrammatic definition of CI encompasses the notions in [20].
5. On the logical side, we define CI via the DIBI formulas applied in Bao et al. [3] to characterise both probabilistic CI and join dependency in the probabilistic and relational models, respectively.
6. We prove that the string diagrammatic definition and the logical formulations of CI coincide on a fragment of these categorical models.

$$
\begin{aligned}
P, Q ::= {} & p \in \mathcal{AP} \mid \top \mid I \mid P \wedge Q \mid P * Q \mid P \fatsemi Q \\
& \mid P * Q \mid P \mathrel{-\!\!*} Q \mid P \fatsemi Q \mid P \multimap Q \mid P \mathrel{\circ\!\!-} Q
\end{aligned}
$$

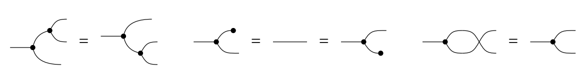
Figure 1. Syntax of DIBI. $\mathcal{AP}$ is a set of atomic propositions.

The general class of DIBI models offers the bridge between the logical and categorical perspectives on CI, which is central to our paper, but they are also of more general interest as DIBI was originally designed to be a logic of assertions for a program logic to reason about (in)dependence. The general class of models offers a platform for the future design of program logics for other models, as the multiset relational one we present later in the paper.

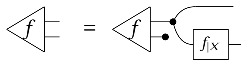
Synopsis. Section 2 is for preliminaries. In Section 3 we propose two notions of conditional independence in the context of copy-discard (CD) categories. Section 4 contains our main contribution: categorical DIBI models and its applications. Missing proofs can be found in Appendix A, and detailed discussion in comparison with Simpson [20] are in Appendix B.

2 Preliminaries

Categories. We assume familiarity with symmetric monoidal categories (SMC) and the graphical representation of their morphisms as string diagrams, see e.g. [19]. In particular, our convention here is to read string diagrams from left to right, with tensor products read from top to bottom. We write $\otimes$ for the tensor product, I for the unit and χ for the symmetry natural transformations — with components $\chi_{A,B} : A \otimes B \to B \otimes A$ called 'swappings' — of an SMC $\mathbf{C}$. We will also rely on the notion of Markov category [10], which we now recall. First, a *CD* category (for **c**opy, **d**iscard) is an SMC $\langle \mathbf{C}, \otimes, I \rangle$ with 'copier' $_C$ and 'discarder' $_C$ morphisms for each object C, which form a commutative comonoid:

 (1)

Moreover, both and are compatible with the monoidal structure. We say *is natural* if = for arbitrary morphism f, and a *Markov category* is a CD category in which the discarder is natural. A Markov category *has conditionals* if for each morphism $f : I \to X \otimes Y$, there exists $f_{|X} : X \to Y$ such that



Monads. An endofunctor $\mathcal{T} : \mathbf{C} \to \mathbf{C}$ is a *monad* if it has a unit $\eta^{\mathcal{T}}$ and a multiplication $\mu^{\mathcal{T}}$ natural transformations satisfying certain compatible conditions. A monad $\mathcal{T}$ on a SMC is *strong* if it has a 'strength' natural transformation with components $\mathrm{st}_{X,Y} : X \otimes \mathcal{T}Y \to \mathcal{T}(X \otimes Y)$ satisfying certain compatibility conditions. It is *commutative* if the two obvious ways of constructing 'double strength' natural transformation with components $\mathrm{dst}_{X,Y} : \mathcal{T}X \otimes \mathcal{T}Y \to \mathcal{T}(X \otimes Y)$ from $\mathrm{st}_{X,Y}$ are identical. $\mathcal{T}$ is *affine* if it preserves the tensor unit up to isomorphism, namely $\mathcal{T}I \cong I$ [15]. Every monad $\mathcal{T} : \mathbf{C} \to \mathbf{C}$ induces a Kleisli category $\mathcal{K}\ell(\mathcal{T})$, and we will write the morphisms in $\mathcal{K}\ell(\mathcal{T})$ as $X \rightarrow\!\!\!\!\!\rightarrow Y$ to distinguish them from their counterpart $X \to \mathcal{T}Y$ in $\mathbf{C}$. In this work we will encounter the discrete probability (resp. subprobability) distribution monad $\mathcal{D}$ (resp. $\mathcal{D}_{\leq}$), non-empty powerset monad $\mathcal{P}_i$, multiset monad $\mathcal{T}_{\mathcal{M}}$ for some semiring $\mathcal{M}$ (see below), as well as their induced Kleisli categories.

Multisets. Fix a monoid $\mathcal{M} = \langle M, +, 0 \rangle$, a *multiset* over a set X is a function $\varphi : X \to M$ such that the *support* $supp(\varphi) := \{x \in X \mid \varphi(x) \neq 0\}$ is finite. Suppose $supp(\varphi) = \{x_1, \ldots, x_n\}$, then we use the 'ket' notation and represent φ as $\varphi(x_1)|x_1\rangle + \cdots + \varphi(x_n)$. The monoid $\mathcal{M}$ induces a *multiset functor* $\mathcal{T}_{\mathcal{M}}$ on $\mathbf{Set}$: on objects, $\mathcal{T}_{\mathcal{M}}(X)$ is the set of all multisets over X; on morphisms, $\mathcal{T}_{\mathcal{M}}(X \xrightarrow{f} Y)$ maps φ to $\sum_{x \in supp(\varphi)} \varphi(x)|f(x)\rangle$. When $\mathcal{M}$ in addition has a semiring structure $\langle M, +, 0, \times, 1 \rangle$, $\mathcal{T}_{\mathcal{M}}$ is a monad.

We will be interested in multiset relations over a set of attributes and their compositions. An *attribute* $\mathbf{a}$ is simply a set of values. Given a finite set of attributes $\mathbf{A} = \{\mathbf{a_1}, \ldots, \mathbf{a_k}\}$, an $\mathbf{A}$-*indexed tuples* is a function $t : \mathbf{A} \to \bigcup_i \mathbf{a_i}$ such that $t(\mathbf{a_i}) \in \mathbf{a_i}$ for all i. The set of all $\mathbf{A}$-indexed tuple is denoted as $\mathbf{Tup}[\mathbf{A}]$. The projection of t to a subset $\mathbf{B} \subseteq \mathbf{A}$ is $t^{\mathbf{B}} : \mathbf{B} \to \bigcup_{\mathbf{a_i} \in \mathbf{B}} \mathbf{a_i}$ such that $t^{\mathbf{B}}(\mathbf{a_i}) = t(\mathbf{a_i})$. An $\mathbf{A_1}$-indexed tuple t_1 and an $\mathbf{A_2}$-indexed tuple t_2 (both $\mathbf{A_i}$ are subsets of $\mathbf{A}$) can be *combined* if ${t_1}^{\mathbf{A_1} \cap \mathbf{A_2}} = {t_2}^{\mathbf{A_1} \cap \mathbf{A_2}}$, and their *combination* $t_1 \bullet t_2 \in \mathbf{Tup}[\mathbf{A_1} \cup \mathbf{A_2}]$ is defined as

$$t_1 \bullet t_2(\mathbf{a_i}) := \begin{cases} t_1(\mathbf{a_i}) & \text{if } \mathbf{a_i} \in \mathbf{A_1} \\ t_2(\mathbf{a_i}) & \text{if } \mathbf{a_i} \in \mathbf{A_2} \end{cases}$$

Given a semiring $\mathcal{M}$, $\mathcal{M}$-*multiset relations* (or simply *multiset relations* when $\mathcal{M}$ is clear from the context) over an attribute set $\mathbf{A}$ are elements of $\mathcal{T}_{\mathcal{M}}(\mathbf{Tup}[\mathbf{A}])$. For two multiset relations $\varphi_1 \in \mathcal{T}_{\mathcal{M}}(\mathbf{Tup}[\mathbf{A_1}])$ and $\varphi_2 \in \mathcal{T}_{\mathcal{M}}(\mathbf{Tup}[\mathbf{A_2}])$ (where each $\mathbf{A_i} \subseteq \mathbf{A}$), their *natural join* $\varphi_1 \bowtie \varphi_2 \in \mathbf{Tup}[\mathbf{A_1} \cup \mathbf{A_2}]$ is defined by $\sum_{t_i \in supp(\varphi_i)} (\varphi_1(t_1) \times \varphi_2(t_2))|t_1 \bullet t_2\rangle$. A multiset relation $\varphi \in \mathbf{Tup}[\mathbf{A_1} \cup \mathbf{A_2}]$ satisfies *join dependency* $\mathbf{A_1} \bowtie \mathbf{A_2}$, if there exist $\varphi_i \in \mathcal{T}_{\mathcal{M}}(\mathbf{Tup}[\mathbf{A_i}])$ for $i = 1, 2$ such that $\varphi = \varphi_1 \bowtie \varphi_2$.

In particular, if we take the semiring $\mathbf{2} = \langle \{0, 1\}, \vee, 0, \wedge, 1 \rangle$, then $\mathbf{2}$-multisets over X are finite subsets of X, and we retrieve the notions of relations over an attribute set and their natural joins. If we take the semiring $\langle [0, 1], +, 0, \times, 1 \rangle$ and put the extra 'normalisation' restriction that $\sum_{x \in supp(\varphi)} \varphi(x) = 1$ (resp. $\sum_{x \in supp(\varphi)} \varphi(x) \leq 1$), then the resulting functor is the distribution monad $\mathcal{D}$ (resp. sub-distribution monad $\mathcal{D}_{\leq}$).

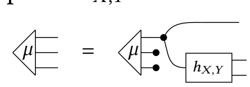
3 Conditional independence for copy-discard categories

In this section we propose two notions of conditional independence in the context of copy-discard (CD) categories, both via string diagrams, which will later to be used for our categorical models in Subsection 4.3. We then show their coincidence when the CD category has richer structure, and illustrate their divergence in non-Markov categories via a concrete example. Subsection 3.1 is a detour which makes a first step towards comparison between the string diagrammatic notions of CI with another categorical approach to CI—the random variable axiomatisation in Simpson [20].

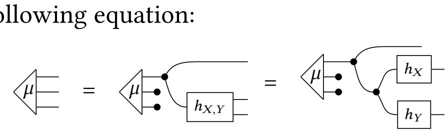
Our abstract definition is motivated by the traditional notion of conditional independence for probabilistic distributions, which we now recall. Let μ be a probabilistic joint distribution on X, Y, Z. X and Y are conditionally independent given Z, denoted as $X \perp\!\!\!\perp_{pr} Y \mid Z$, if for arbitrary values x, y, z for X, Y, Z such that $\mu(Z = z) \neq 0$,

$$\mu(X = x, Y = y \mid Z = z) = \mu(X = x \mid Z = z) \cdot \mu(Y = y \mid Z = z) \tag{2}$$

where $\mu(\cdot \mid \cdot)$ is a well-defined conditional distribution for μ. This scenario can be smoothly formulated using a categorical approach. Consider the SMC $\langle \mathcal{K}\ell(\mathcal{D}), \times, \mathbf{1} \rangle$ whose morphisms $X \rightarrow\!\!\!\!\!\rightarrow Y$ are functions mapping each $x \in X$ to a finite probability distribution on Y. Note that distributions on a set X are in one-one correspondence with $\mathcal{K}\ell(\mathcal{D})$-morphisms of the type $\mathbf{1} \rightarrow\!\!\!\!\!\rightarrow X$, thus the latter are also referred to as 'distributions'. Now, given a joint distribution $\mu : \mathbf{1} \rightarrow\!\!\!\!\!\rightarrow X \times Y \times Z$, a conditional distribution $\mu(X, Y \mid Z)$ is some $\mathcal{K}\ell(\mathcal{D})$-morphism $h_{X,Y} : X \times Y \rightarrow\!\!\!\!\!\rightarrow Z$ such that



where -< and -• form the CD structure in $\mathcal{K}\ell(\mathcal{D})$ (see Section 2). Then the conditional independence (2) is captured by the following equation:



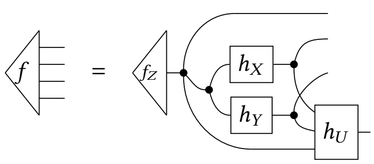
(3)

where $h_X : Z \rightarrow\!\!\!\!\!\rightarrow X$ and $h_Y : Z \rightarrow\!\!\!\!\!\rightarrow Y$ represent conditional probability distributions $\mu(X \mid Z)$ and $\mu(Y \mid Z)$, respectively.

A diagrammatic presentation such as the one in (3) has the benefit of highlighting in a precise way how the variables involved in the conditional independence statement interact with each other. Moreover, because categories others than $\mathcal{K}\ell(\mathcal{D})$ have the structure to meaningfully interpret equation (3), it naturally suggests generalisation beyond the probabilistic setting, as explored in [5, 10]. In those works, the most general setting considered is the one of Markov categories (*cf.* Section 2). However, we argue that one may study CI with even less assumptions, namely the setting of CD categories (also recalled in Section 2). By moving from Markov to CD categories, we drop the naturality condition on $\multimap\!\bullet$, allowing to consider significant case studies such as subprobability distributions.

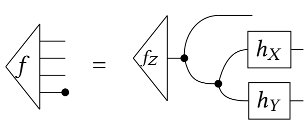
By analogy with the probabilistic case $\mathcal{K}\ell(\mathcal{D})$, we call *distributions* the morphism of type $I \to X$ in a CD category, and call them *joint distribution* in case we want to emphasise that X is of the form $Y \otimes Z$. Let us fix a CD category $\langle \mathbf{C}, \otimes, I, \asymp \rangle$ and a distribution $f : I \to Z \otimes X \otimes Y \otimes U$ in $\mathbf{C}$, where Z, X, Y, U are pairwise disjoint tensor products of distinct objects. We first define the two notions of CI and then explain the intuition behind them.

Definition 3.1 (Process conditional independence)**.** Distribution f satisfies that X and Y are *process independent conditioned on* Z , denoted as $X \perp\!\!\!\perp_P Y \mid Z$, if there exist $f_Z : I \to Z$, $h_X : Z \to X \otimes X'$, $h_Y : Z \to Y \otimes Y'$, and $h_U : X \otimes X' \otimes Y \otimes Y' \otimes Z \to U'$ such that $U = X' \otimes Y' \otimes U'$ and



(4)

Definition 3.2 (Markov conditional independence)**.** Distribution f satisfies that X and Y are *Markov independent conditioned on* Z, denoted as $X \perp\!\!\!\perp_M Y \mid Z$, if there exist $f_Z : I \to Z$, $h_X : Z \to X$, $h_Y : Z \to Y$ such that



(5)

When Z is empty, we omit it and write $X \perp\!\!\!\perp_P Y$ and $X \perp\!\!\!\perp_M Y$ for process independence and Markov independence, respectively.

The two definitions generalise 2 in two ways. First, by being phrased in the abstract setting of CD categories. Second, by dealing with those variables in the joint distribution which do not appear in the CI statement of interest. This second generalisation accounts for the scenario in which one is interested only about the CI property for a proper subset of the variables, see e.g. Example 3.5. It turns out to be essential when comparing our notions to the expressivity of DIBI logic formulas, as we will see in Section 4.

The two definitions take a different approach to the handling of these 'extra' variables, as we now illustrate. In both cases, X, Y, Z are the objects appearing in the CI statement, and U consists of those objects in distribution f which are 'irrelevant' to the CI statement of interest. Both definitions capture the independence of X and Y via having two boxes h_X and h_Y side-by-side. Note first that when f does not have irrelevant object U — indeed the context in which string diagrammatic notions of CI is mostly discussed for example in [10] — then (4) and (5) are exactly the same diagram. Otherwise, they deal with the irrelevant objects in different manners. In Definition 3.1 one retains U, in order to process it via the morphism h_U — the name 'process' refers to the role played by h_U in the motivating example below (Example (3.3)). On the other hand, in Definition 3.2, one simply discards U via $\multimap\!\bullet_U$. The first notion is our main definition and is to be used later in Section 4. It is motivated by probabilistic programs like Example 3.3 which constructs a subprobability distribution over possible valuations of variables. There this notion of process CI captures the intuition that certain variables are defined independently given the value of some other variable. The second notion is a reasonable treatment of irrelevant variables in Markov categories (thus the name 'Markov CI') because of the naturality of $\multimap\!\bullet$, and we state it for CD categories to facilitate comparison between these two approaches.

Example 3.3. We motivate the process CI with an example. Consider the program $\mathbb{P}$ in Figure 2a, where z, x, y, u are Boolean variables taking values in $\mathbf{2} = \{0, 1\}$. First, flip a fair coin and store the result in z. If $z = 0$, flip two coins: first a fair coin, second a coin with probability $\frac{1}{2}$ as 0, $\frac{1}{3}$ as 1 (and $\frac{1}{6}$ to drop on the floor thus abort). Store the results in x and y. Otherwise flip a coin with biase $\frac{1}{4}$ twice and store the results in x and y. Finally, assigns 0 to u unless $x = y = 1$, in which case abort. This program calcuates a subdistribution μ whose non-zero entries are listed in Figure 2b.

$z \xleftarrow{\$} \frac{1}{2}|0\rangle + \frac{1}{2}|1\rangle$
if $z = 0$ **then**
  $x \xleftarrow{\$} \frac{1}{2}|0\rangle + \frac{1}{2}|1\rangle$
  $y \xleftarrow{\$} \frac{1}{2}|0\rangle + \frac{1}{3}|1\rangle$
else
  $x \xleftarrow{\$} \frac{1}{4}|0\rangle + \frac{3}{4}|1\rangle$
  $y \xleftarrow{\$} \frac{1}{4}|0\rangle + \frac{3}{4}|1\rangle$
end if
if not $x = y = 1$ **then**
  $u \leftarrow 0$
else
  abort
end if

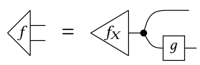
(a) Probabilistic program $\mathbb{P}$

z	x	y	u	μ
0	0	0	0	$^1/_8$
0	0	1	0	$^1/_{12}$
0	1	0	0	$^1/_8$
1	0	0	0	$^1/_{32}$
1	0	1	0	$^3/_{32}$
1	1	0	0	$^3/_{32}$

(b) Subprobabillity distribution μ

Figure 2. From probabilistic programs to distributions

Moreover, these two notions of CI coincide when $\mathbf{C}$ has rich enough structure. Recall that a CD category *has conditionals* if for every distribution $f: I \to X \otimes Y$, there exists f_X and g such that



Proposition 3.4. *Suppose* $\mathbf{C}$ *is a Markov category with conditionals, then for any distribution* $f: I \to Z \otimes X \otimes Y \otimes U$, $X \perp\!\!\!\perp_P Y \mid Z$ *if and only if* $X \perp\!\!\!\perp_M Y \mid Z$

Proposition 3.4 justifies why these two notions are not separated in standard literature on CI in Markov categories: they coincide in 'well-behaved' categories such as $\mathcal{K}\ell(\mathcal{D})$ and $\mathcal{K}\ell(\mathcal{P}_i)$. But in the general setting these two notions diverge in non-Markov CD categories, as illustrated with the following example.

Example 3.5. We know that $\mathcal{K}\ell(\mathcal{D}_{\leq})$ (see Section 2) is a non-Markov CD category. Consider a distribution in $\mathcal{K}\ell(\mathcal{D}_{\leq})$ slighted modified from Example 3.3, say $f: \mathbf{1} \to X \otimes Y \otimes U$ with $X = \{0_x, 1_x\}$, $Y = \{0_y, 1_y\}$, $U = \{0_u, 1_u\}$, which represents the subdistribution $\frac{1}{4}|0_x 0_y 0_u\rangle + \frac{1}{6}|0_x 1_y 0_u\rangle + \frac{1}{4}|1_x 0_y 1_u\rangle$. We claim that f satisfies $X \perp\!\!\!\perp_P Y$ but not $X \perp\!\!\!\perp_M Y$. For the former, let $h_X: \mathbf{1} \to X$ and $h_Y: \mathbf{1} \to Y$ stand respectively for the sub-distributions $\frac{1}{2}|0_x\rangle + \frac{1}{2}|1_x\rangle$ and $\frac{1}{2}|0_y\rangle + \frac{1}{3}|1_y\rangle$; $h_U: X \times Y \to U$ maps $(0_x, 0_y), (0_x, 1_y), (1_x, 0_y)$ to $1|0_u\rangle$ and $(1_x, 1_y)$ to the empty distribution $\varnothing$. Then one can verify that (4) holds for such f, h_X, h_Y and h_U (f_Z is trivially id_1). For the latter, post-composing $-\!\bullet_U$ to f returns the subprobability distribution $\frac{1}{4}|0_x 0_y\rangle + \frac{1}{6}|0_x 1_y\rangle + \frac{1}{4}|1_x 0_y\rangle$ on $X \times Y$, which cannot be the product of two subprobability distributions on X and Y respectively.

Finally, we slightly generalise these two notions of CI to allow overlapping objects. Currently we only allow CI statements of the form $X \perp\!\!\!\perp Y \mid Z$ where the overlap of X and Y lies in Z.

Definition 3.6. Following the assumption for Definition 3.1 and 3.2, assume further that $Z = Z_0 \otimes Z_1$, and let $X' = X \otimes Z_0$, $Y' = Y \otimes Z_0$. Then define that f satisfies $X' \blacklozenge Y' \mid Z$ if and only if f satisfies $X \blacklozenge Y \mid Z$, for $\blacklozenge \in \{\perp\!\!\!\perp_P, \perp\!\!\!\perp_M\}$.

In Definition 3.6, the CI condition of objects X' and Y' given Z is reduced to their disjoint counterpart X and Y. Intuitively this is because the Z_0 part in both X' and Y' is already determined by Z, thus does not affect whether X' and Y' are conditional independent on Z.

In a nutshell, the novelty of our definitions lies in three aspects. First, we assume less structure of the underlying categories, thus the definitions fit for a more general class of categories. Second, we consider distributions whose objects may contain variables irrelevant to the CI statement. Third, we allow overlap of objects X and Y in $X \perp\!\!\!\perp Y \mid Z$ — though in this paper their overlap is restricted to be contained in Z[1]. Later on we will see how this categorical notion of CI (Definition 3.1) fits into the categorical DIBI models (to be introduced in Subsection 4.1).

3.1 Comparison with (local) independence structure

In this subsection we take a detour to compare the string diagrammatic approaches to CI with the categorical approach in Simpson [20]. Simpson considers conditional independence on abstract random variables. Instead of axiomatising CI, Simpson proposes so-called (local) independence structure in terms of multispans that should be satisfied by any reasonable definition of CI for random variables (RV), and demonstrates his framework in various settings, including CI in probability theory, database theory, and nominal sets. We refer to Simpson's axiomatisation approach as MS-CI, standing for 'multispan'. We propose a string diagrammatic notion of CI for general random variables. This notion covers several examples in Simpson [20], and has some general connections with MS-CI.

Here we demonstrate our string diagrammatic definition for random variable CI in the concrete setting of discrete probability distributions, and defer the details to Appendix B.

Recall the Markov category $\langle \mathcal{K}\ell(\mathcal{D}), \times, \mathbf{1}\rangle$ of discrete probability distributions. The coslice category $I/\mathcal{K}\ell(\mathcal{D})$ has as objects (X, ω_X) where ω_X is a morphism $I \to X$. A *random variable* is defined as a deterministic morphism (see Definition B.1) $f: (X, \omega_X) \to (Y, \omega_Y)$ in $I/\mathcal{K}\ell(\mathcal{D})$, thus a function $f: X \to \mathcal{D}(Y)$ such that (i) $\omega_Y = f \circ \omega_X$, (ii) for all $x \in X$, $f(x) = 1|y\rangle$ for some $y \in Y$. We denote the category of random variables in $I/\mathcal{K}\ell(\mathcal{D})$ as $\mathrm{RV}(I/\mathcal{K}\ell(\mathcal{D}))$.

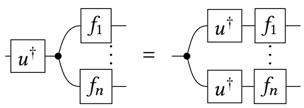
Conditional independence given some (U, ω_U) is defined in the slice category $(U, \omega_U)\backslash\mathrm{RV}(I/\mathcal{K}\ell(\mathcal{D}))$. A set of random variables $\{f_i: (X, \omega_X) \to (Y_i, \omega_{Y_i})\}_{i=1}^n$ in this slice category means the following setting of commutative triangles for $i = 1, \dots, n$:

$$\begin{tikzcd}
(X, \omega_X) \arrow[rr, "f_i"] & & (Y_i, \omega_{Y_i}) \\
& (U, \omega_U) \arrow[ul, "u"] \arrow[ur, "v_i"'] &
\end{tikzcd} \tag{6}$$

We further restrict ourselves to a subcategory where every morphism has a unique *Bayesian inversion* [6, 10]: recall that a Bayesian inversion of $f: (X, \omega_X) \to (Y, \omega_Y)$ is a morphism $f^\dagger: (Y, \omega_Y) \to (X, \omega_X)$ such that $f \circ f^\dagger = id$. A morphism $f: (X, \omega_X) \to (Y, \omega_Y)$ has a unique Bayesian inversion if and only if for all $y \in Y$, there exists $x \in X$ such that $f(x) = 1|y\rangle$. We denote the resulting category by $\mathbf{D}$. Intuitively, RVs $f_1, \dots, f_n$ are conditionally independent given (U, ω_U) if given the 'feedback' $u^\dagger: (U, \omega_U) \to (X, \omega_X)$, the outputs for $f_1, \dots, f_n$ are independent.

[1] In fact, one may loosen this restriction and allow for arbitrary intersection of X and Y. We do not take this path for the sake of simplicity, as it is not needed for our developments.

Then we define RVs $f_1, \dots, f_n$ to be conditionally independent given (U, ω_U) if



(7)

In our setting of probability distributions, this equality of diagrams boils down to the following equation for all $z \in U$ and $y_i \in Y_i$ satisfying $v_i(y_i) = 1|z\rangle$:

$$\sum_{\forall i=1,\dots,n, f_i(x)=y} \frac{\mu_X(x)}{\mu_U(z)} = \prod_{i=1,\dots,n} \frac{\mu_{Y_i}(y_i)}{\mu_U(z)} \qquad (8)$$

where $\mu_X := \omega_X(*)$ is the distribution over X represented by ω_X, and similarly for μ_U, μ_{Y_i}. Equation (8) is exactly the notion of conditional independence for random variables in Example 5.1 from Simpson [20]. Moreover, one can retrieve plain independence from conditional independence in the current setting simply by taking U to be the singleton set $\mathbf{1}$. Intuitively this makes sense since plain independence means conditional independence given a trivial observation. The resulting equivalence states precisely the independence for random variables in Example 2.1 from Simpson [20].

4 Categorical DIBI models

So far, we have defined two notions of conditional independence in terms of string diagrams for CD categories. As an application, we show how CD categories can support models of DIBI, a recently-proposed logic for reasoning about conditional independence [3], and adapt the string diagrammatic definitions of CI from Section 3 to DIBI models. Like other bunched logics, DIBI is a substructural logic that augments standard propositional logic with additional conjunctions and their adjoint implications. More specifically, DIBI features a commutative conjunction $P * Q$, intuitively stating that P and Q are independent, and a non-commutative conjunction $P \fatsemi Q$, intuitively stating that Q may depend on P.

Bao, et al. [3] propose two concrete models of DIBI and show that DIBI assertions can capture conditional independence in relations and probability distributions. The states in their concrete models can be seen as Kleisli arrows for the non-empty powerset monad and the discrete distribution monad, respectively. Though they use different monads, these two models share some common patterns in their definition of operations on states and enjoy similar properties. For instance, the same formula asserts conditional independence notions in two models – join dependency in relations and probabilistic conditional independence in discrete distributions. One natural question is whether other notions of conditional independence (e.g., those in subdistributions, multisets, and Hilbert spaces) can be captured in other concrete models of DIBI. However, constructing concrete models of DIBI and checking the frame conditions requires a significant amount of tedious computations.

In this section we use the language of string diagrams to define general, categorical models of DIBI, generalizing the known concrete models. First, in Section 4.1 we show that every CD category induces a DIBI frame (Theorem 4.8). In particular, we can define states as 'input-preserving kernels' (Definition 4.2), and define $\sqsubseteq$ (Definition 4.3), $\oplus$, $\odot$ (Definition 4.6) as relations and operations on the string diagrammatic presentations of these kernels. We demonstrate the power of this general construction by providing three concrete DIBI frames in Subsection 4.2, two from the original DIBI paper [3], and a new model based on the multiset monad. Besides generalising the class of DIBI models, our approach allows the rather involved and laborious computations in prior constructions of concrete DIBI frames to be replaced with cleaner and simpler reasoning in terms of string diagrams. Finally, we define two notions of CI for the categorical DIBI models. The first one is an adaption of Definition 3.1, and the second one is via certain DIBI formulas. We show these two notions coincide.

4.1 Categorical DIBI models: theory

In this subsection we fix an arbitrary CD category $\mathbf{C}$ and work towards defining categorical DIBI frames and models based on $\mathbf{C}$ (Theorem 4.8 and Corollary 4.11). In particular, the state space of such categorical DIBI frames consists of so-called 'kernels' (Definition 4.2) which are categorical abstraction of programs that copy the input to the output. Kernels and operations on them will be presented via string diagrams. String diagrams simplifies the definition of the operations on kernels and provide a convenient setting in which to check that kernels satisfy the requirements to form a model of DIBI. These conditions are also known as the *frame conditions* for DIBI from Bao et al. [3].

Definition 4.1. A DIBI frame is a tuple $\mathcal{X} = (X, \sqsubseteq, \oplus, \odot, E)$, where $\sqsubseteq$ is a preorder on X, $E \subseteq X$, and $\oplus, \odot \colon X \to \mathcal{P}(X)$ are binary operations, satisfying the frame conditions in Figure 3. A DIBI model is a DIBI frame together with a valuation function $\mathcal{V} \colon \mathcal{AP} \to X$ mapping each atomic proposition to a set of states.

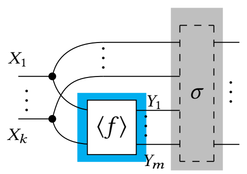
We refer to X as the set of *states*. Intuitively, the frame conditions ensure that $\odot$ and $\oplus$ behave like abstract versions of sequential and parallel compositions, and the pre-order $\sqsupseteq$ describes when a state can be extended to a larger state. In the concrete DIBI frames [3], the states are a special class of Kleisli arrows (also known as *kernels*) for the powerset and distribution monads, where the input is copied to the output. We formalise this notion of kernels in arbitrary CD categories.

Definition 4.2 (Input-preserving kernels)**.** A $\mathbf{C}$-morphism f is called an *input-preserving* $\mathbf{C}$*-kernel*, or simply $\mathbf{C}$-kernel,

($\oplus$ Down-Closed)	$z \in x \oplus y \wedge x \sqsupseteq x' \wedge y \sqsupseteq y'$	$\Longrightarrow$	$\exists z'(z' \in x' \oplus y' \wedge z \sqsupseteq z')$
($\odot$ Up-Closed)	$z \in x \odot y \wedge z' \sqsupseteq z$	$\Longrightarrow$	$\exists x', y'(x' \sqsupseteq x \wedge y' \sqsupseteq y \wedge z' \in x' \odot y')$
($\oplus$ Commutativity)	$z \in x \oplus y$	$\Longrightarrow$	$z \in y \oplus x$;
($\oplus$ Associativity)	$w \in t \oplus z \wedge t \in x \oplus y$	$\Longrightarrow$	$\exists s(s \in y \oplus z \wedge w \in x \oplus s)$;
($\oplus$ Unit Existence)	$\exists e \in E(x \in e \oplus x)$;		
($\oplus$ Unit Coherence)	$e \in E \wedge x \in y \oplus e$	$\Longrightarrow$	$x \sqsupseteq y$;
($\odot$ Associativity)	$\exists t(w \in t \odot z \wedge t \in x \odot y)$	$\Longrightarrow$	$\exists s(s \in y \odot z \wedge w \in x \odot s)$;
($\odot$ Unit Existence$_{\mathrm{L}}$)	$\exists e \in E(x \in e \odot x)$;		
($\odot$ Unit Existence$_{\mathrm{R}}$)	$\exists e \in E(x \in x \odot e)$;		
($\odot$ Coherence$_{\mathrm{R}}$)	$e \in E \wedge x \in y \odot e$	$\Longrightarrow$	$x \sqsupseteq y$;
(Unit Closure)	$e \in E \wedge e' \sqsupseteq e$	$\Longrightarrow$	$e' \in E$;
(Reverse Exchange)	$x \in y \oplus z \wedge y \in y_1 \odot y_2 \wedge z \in z_1 \odot z_2$	$\Longrightarrow$	$\exists u, v(u \in y_1 \oplus z_1 \wedge v \in y_2 \oplus z_2 \wedge x \in u \odot v)$.

Figure 3. Conditions for DIBI frames (outermost universal quantification omitted for readability).

if it is of the shape

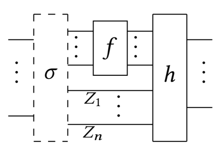


(9)

where $X_1, \ldots, X_k, Y_1, \cdots, Y_m$ are distinct **C**-objects, $\langle f \rangle$ is a **C**-morphism, σ consists of solely $\times$s from the symmetric structure of **C** to change the order of the output wires $X_1, \ldots, X_k, Y_1, \ldots, Y_m$.

When necessary, we use blue and grey box to indicate the $\langle f \rangle$ and σ part in **C**-kernel f respectively, as in (9). We denote the set of **C**-kernels as $\mathrm{Ker}(\mathbf{C})$. Next we define the order relation $\sqsubseteq$ on **C**-kernels.

Definition 4.3. Given two input-perserving **C**-kernels f and g, we write $f \sqsubseteq g$ if there exists **C**-objects $Z_1, \cdots, Z_n$ and an input-preserving **C**-kernel h such that

$g =$ 

(10)

where σ consists solely of $\times$s that change the order of the input wires to match the domains of g and $f \otimes id_{Z_1 \otimes \cdots \otimes Z_n}$.

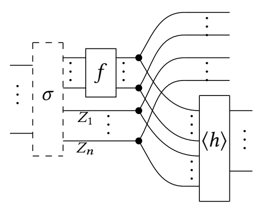
Intuitively, $f \sqsubseteq g$ means that g is an extension of f in a canonical way: first compose f in parallel with some additional objects, then extend the codomain with some other kernel.

Proposition 4.4. *$\sqsubseteq$ is a pre-order.*

Informally, one cannot always expect to retrieve a morphism from its extension (10), since the latter may contain extra information that destroy the structure of the original morphism. This is possible when the underlying category has some good properties that enable one to forget the extra information safely.

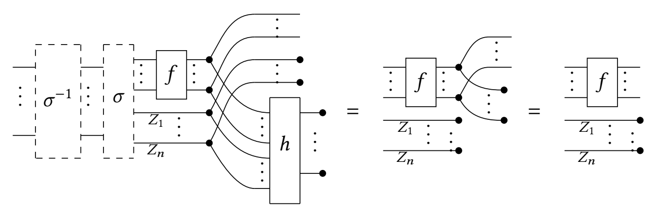
Proposition 4.5. *Suppose* **C** *is a Markov category together with a unit $\bullet\!-_C$ for each object* **C**, *that is compatible with discarder $-\!\bullet$: $-\!\bullet_C \circ \bullet\!-_C = id_I$. If f, g are* **C**-*kernels such that $f \sqsubseteq g$, then given g and the type of f one can uniquely determine f by composition: $f = h_{post} \circ g \circ h_{pre}$ for some* **C**-*morphisms h_{pre}, h_{post}.*

Proof. $f \sqsubseteq g$ means that there exists h such that g can be decomposed as in (10). More precisely, since h is a **C**-kernel, (10) becomes the following form:

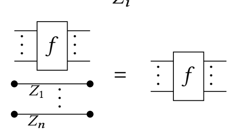


(11)

Then to determine f, it suffices to first pre-compose to (11) the inverse of σ, say σ^{-1}, and after that discard all output wires that are not in the codomain of f:



Finally we get rid of those $-\!\bullet_{Z_i}$ above via $\bullet\!-$:

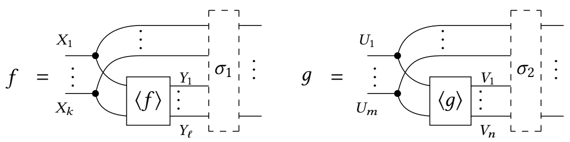


It is easy to see that such f is unique. Suppose f' has the same domain and codomain as that of f, and $f' \sqsubseteq g$, then one can discard all output wires in the left-hand-side of (4.1) and get $f \otimes id_{Z_1 \otimes Z_n} = f' \otimes id_{Z_1 \otimes Z_n}$, which implies that $f = f'$. ◻

Next, we define the parallel and sequential compositions of **C**-kernels. Before that we fix some notations. Given a monoidal product $X_1 \otimes \cdots \otimes X_n$ of distinct objects, suppose

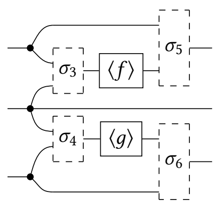
$\{Y_1, \ldots, Y_k\}$ is a subset of $\{X_1, \ldots, X_n\}$, then we use $X_1 \otimes \cdots \otimes X_n - \{Y_1, \ldots, Y_k\}$ to denote the tensor product of objects in $\{X_1, \ldots, X_n\} \setminus \{Y_1, \ldots, Y_k\}$ while keeping the order of objects as in $X_1 \otimes \cdots \otimes X_n$. In particular, $X_1 \otimes \cdots \otimes X_n - \{X_1, \ldots, X_n\}$ is I. For instance, $X_1 \otimes X_2 \otimes X_3 \otimes X_4 - \{X_2, X_4\} = X_1 \otimes X_3$.

Definition 4.6 (Kernel compositions)**.** Given two **C**-kernels f and g of the form



(12)

their *sequential composition* $f \odot g$ is defined if $cod(f) = dom(g)$, and $f \odot g$ is $g \circ f$ in **C**. Their *parallel composition* $f \oplus g$ is of the following form when defined



(13)

and it is defined if $\{X_1, \ldots, X_k\} \cap \{U_1, \ldots, U_m\} = \{Y_1, \ldots, Y_\ell\} \cap \{V_1, \ldots, V_n\}$ (this intersection denoted as $\{Z_1, \ldots, Z_i\}$), where $f\colon X_1 \otimes \cdots \otimes X_k \to Y_1 \otimes \cdots \otimes Y_\ell$ and $g\colon U_1 \otimes \cdots \otimes U_m \to V_1 \otimes \cdots \otimes V_n$. The domain and codomain of $f \oplus g$ are respectively:

domain $(X_1 \otimes \cdots \otimes X_k - \{Z_1, \ldots, Z_i\}) \otimes (Z_1 \otimes \cdots \otimes Z_i) \otimes (Y_1 \otimes \cdots \otimes Y_\ell - \{Z_1, \cdots, Z_i\})$

codomain $(U_1 \otimes \cdots \otimes U_m - \{Z_1, \ldots, Z_i\}) \otimes (Z_1 \otimes \cdots \otimes Z_i) \otimes (V_1 \otimes \cdots \otimes V_n - \{Z_1, \cdots, Z_i\})$

where σ_3 and σ_4 respectively change the input wires order to $X_1 \otimes \cdots \otimes X_k$ and $Y_1 \otimes \cdots \otimes U_m$, σ_5 and σ_6 respectively changes the output wires order to $Y_1 \otimes \cdots \otimes Y_\ell - \{Z_1, \cdots, Z_i\}$ and $V_1 \otimes \cdots \otimes V_n - \{Z_1, \cdots, Z_i\}$.

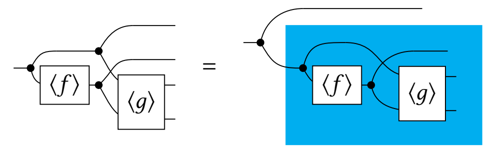
While $\odot$ is standard, the definition of $\oplus$ is less straightforward, for which we give some intuition. Informally, if we think of the (co)domain of a kernel — a tensor product of objects — as a set of variables, then two kernels f and g admit parallel composition if their overlapping variables in the domains coincide with their overlapping variables in the codomains. Then this overlap will be used in three places (thus three copies are generated): as an input to $\langle f \rangle$; as an input to $\langle g \rangle$; as an output of the whole kernel.

The set $\mathsf{Ker}(\mathbf{C})$ of **C**-kernels is closed under both sequential $\odot$ and parallel compositions $\oplus$.

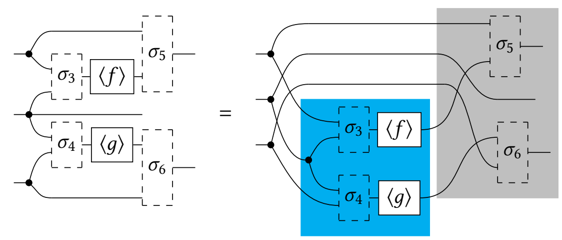
Proposition 4.7. *We follow the assumption of f, g in Definition 4.6.*

1. *If $f \odot g$ is defined, then $f \odot g$ is a **C**-kernel.*
2. *If $f \oplus g$ is defined, then $f \oplus g$ is a **C**-kernel.*

Proof.
1. Suppose $f \odot g$ is defined, then spelling out the definition of **C**-kernels, $f \odot g$ is



thus is also a **C**-kernel.
2. The parallel composition $f \oplus g$ in (13) can be re-arranged as follows:



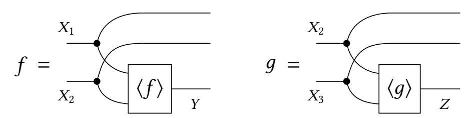
where the grey part only consists of swappings, thus the whole diagram a **C**-kernel.

□

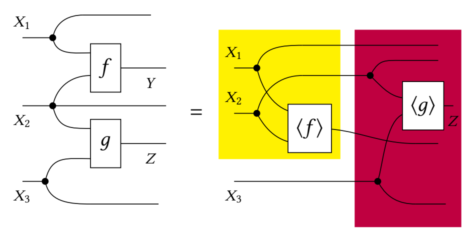
Then **C**-kernels form a DIBI frame.

Theorem 4.8. *$\langle \mathsf{Ker}(\mathbf{C}), \sqsubseteq, \oplus, \odot, \mathsf{Ker}(\mathbf{C}) \rangle$ is a DIBI frame.*

Proof sketch. It suffices to prove the frame conditions in 4.1. We prove $\oplus$ Unit Coherence: given **C**-kernels f, g, if $f \oplus g$ is defined, then $f \sqsubseteq f \oplus g$. Suppose f and g are of the following form



where tensor products X_1 and X_3 are disjoint, Y and Z are disjoint, in the sense that they do not share any same object in the tensor products. Then $f \oplus g$ is well-defined, and $f \oplus g$ is



(14)

In (14), the yellow part is f, the red part is also a **C**-kernel, therefore the diagram on the right-hand-side of (14) witnesses $f \sqsubseteq f \oplus g$. □

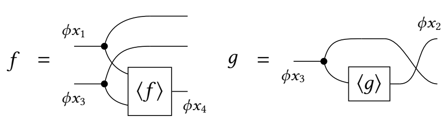
While the frames we have considered so far are a clean class of models for DIBI, it can be difficult to interpret atomic formulas of DIBI in these models because they are quite abstract. We give a slightly more concrete variant of our categorical DIBI frames which support models for DIBI with atomic formulas that mention variables (see Subsection 4.3). Such atomic propositions can be useful for more concrete applications; for instance, they enable DIBI to be used as an assertion logic for reasoning about probabilistic programs with mutable variables [3].

Suppose *Var* is a set of variables that is linearly ordered by $\preceq$, say $x_0 \preceq x_1 \preceq x_2 \preceq \dots$. Also, we fix an arbitrary injective function $\phi\colon Var \to \mathbf{Ob}(\mathbf{C})$ that selects one distinct object for each variable. This enables one to uniquely determine a monoidal product over a set of distinct objects in the image of ϕ, whose order is decided by $\preceq$. For example, given $x_0 \preceq x_1 \preceq x_2$, the set $\{\phi x_0, \phi x_1, \phi x_2\}$ uniqely determines the monoidal product $\phi x_0 \otimes \phi x_1 \otimes \phi x_2$. With a slight abuse of notation, we write $\phi(x) \preceq \phi(y)$ when $x \preceq y$. We say a monoidal product $X_1 \otimes \cdots \otimes X_k$ *obeys* $\preceq$ if $X_1, \dots, X_k$ are all objects in the image of ϕ, and $X_1 \preceq \cdots \preceq X_k$. Then our new class of categorical frames restricts the input-preserving $\mathbf{C}$-kernels to those whose domains and codomains contain only objects in the image of ϕ, and obey the order $\preceq$ on *Var*.

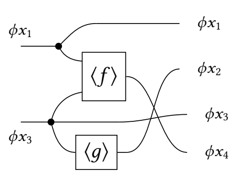
Definition 4.9. A $(\mathbf{C}, \preceq)$-*kernel* f is a $\mathbf{C}$-kernel whose domain and codomain both obey $\preceq$.

We denote the set of $(\mathbf{C}, \preceq)$-kernels as $\mathsf{Ker}_{\preceq}(\mathbf{C})$. We can define $\sqsubseteq$ and $\odot$ on $(\mathbf{C}, \preceq)$-kernels as that for $\mathbf{C}$-kernels. For the parallel composition $\oplus$, we need extra swappings to guarantee that both the domain and codomain obey $\preceq$, as illustrated by the following example.

Example 4.10. Suppose $x_1 \preceq x_2 \preceq x_3 \preceq x_4$, and consider the two kernels $f\colon \phi x_1 \otimes \phi x_2 \to \phi x_1 \otimes \phi x_2 \otimes \phi x_3$ and $g\colon \phi x_2 \to \phi x_2 \otimes \phi x_3$ depicted as follows:



Their parallel composition $f \oplus g$ is



Corollary 4.11. *$\langle \mathsf{Ker}_{\preceq}(\mathbf{C}), \sqsubseteq, \oplus, \odot, \mathsf{Ker}_{\preceq}(\mathbf{C})\rangle$ is a DIBI frame.*

Proof sketch. It suffices to prove $\oplus$ and $\odot$ preserve the order of input and output wires, thus $\mathsf{Ker}_{\preceq}(\mathbf{C})$ is closed under $\oplus$ and $\odot$. Then the verification of the frame conditions follows from the proof of Theorem 4.8. □

4.2 Categorical DIBI models: examples

We will provide three DIBI frames based on the categorical DIBI frames in Subsection 4.1. The first two slight generalises the probabilistic and relational models in [3] by allowing different variables to have different value spaces. The third one is a new multiset frame based on an arbitrary semiring.

For all the examples below, our category $\mathbf{C}$ will be $\mathcal{K}\ell(\mathcal{T})$ for some suitable monad $\mathcal{T}$ on $\mathbf{Set}$. We fix a set of linearly ordered variables $\langle Var, \preceq\rangle$ and $\phi\colon Var \to \mathbf{Ob}(\mathcal{K}\ell(\mathcal{T}))$ such that the images of ϕ are all attributes (see Section 2). With a bit abuse of notation, given a set $S = \{s_1, \dots, s_n\}$ of variables with $s_1 \preceq \cdots \preceq s_n$, we shall use ϕS to denote both the *set* $\phi[S] = \{\phi s \mid s \in S\}$ and the $\mathcal{K}\ell(\mathcal{T})$-*object* $\phi s_1 \otimes \cdots \otimes \phi s_n$ (note the order of this tensor product). When $\mathcal{K}\ell(\mathcal{T})$ is a CD category, Corollary 4.11 tells us that $\mathsf{Ker}_{\preceq}(\mathcal{K}\ell(\mathcal{T}))$ forms a DIBI frame. Moreover, we can regard $(\mathcal{K}\ell(\mathcal{T}), \preceq)$-kernels as functions of type $\mathbf{Tup}[\phi S] \to \mathcal{T}(\mathbf{Tup}[\phi T])$ (satisfying input-preserving conditions), where $S, T \subseteq_f Var$, and $\mathbf{Tup}[\phi S]$ is the set of all ϕS-indexed tuples (see Section 2). This is due to two isomorphisms. First, for a fixed finite set of variables $S = \{s_1, \dots, s_k\}$, the set $\mathbf{Tup}[\phi S]$ is isomorphic to the product $\prod_{i=1}^n \phi s_i$. Second, there is a bijection between the set $\{\mathbf{Tup}[\phi S] \mid S \subseteq_f Var\}$ and the set of objects that occurs as domains or codomains of $(\mathcal{K}\ell(\mathcal{T}), \preceq)$-kernels: there are no two tensor products with the same set of components but differ in the order of tensor product, such as $X_1 \otimes X_2 \otimes X_3$ and $X_3 \otimes X_1 \otimes X_2$. Below we will set $\mathcal{T}$ to be $\mathcal{D}_{\leq}$ (and $\mathcal{D}$), $\mathcal{P}_i$ and $\mathcal{T}_{\mathcal{M}}$. For monad $\mathcal{D}$ and $\mathcal{P}_i$, the fact that $\mathcal{K}\ell(\mathcal{T})$ is a CD category (indeed Markov, but we do not need the property for the construction of the DIBI frame) follows immediately from the following lemma.

Lemma 4.12 ([10]). *Let $\langle \mathbf{C}, \otimes, I\rangle$ be a Markov category, and $\mathcal{T}$ an affine commutative strong monad on $\mathbf{C}$ (see Section 2). Then the Kleisli category $\mathcal{K}\ell(\mathcal{T})$ is again a Markov category in a canonical way.*

Subprobabilistic frame. Let $\mathcal{T}$ be the subprobability distribution monad $\mathcal{D}_{\leq}$, then $(\mathcal{K}\ell(\mathcal{T}), \preceq)$ kernels are exactly functions $f\colon \mathbf{Tup}[\phi S] \to \mathcal{D}_{\leq}(\mathbf{Tup}[\phi T])$ for some finite subsets $S \subseteq T$ of *Var* such that for arbitrary $m \in \mathbf{Tup}[\phi S]$ and $n \in supp(f(m))$, $m(\phi s) = n(\phi s)$ for all attributes $\phi s \in \phi S$.

Kernels $f\colon \mathbf{Tup}[\phi S] \to \mathcal{D}_{\leq}(\mathbf{Tup}[\phi T])$ and $g\colon \mathbf{Tup}[\phi T] \to \mathcal{D}_{\leq}(\mathbf{Tup}[\phi U])$ can be composed sequentially using Kleisli composition. Two kernels $f\colon \mathbf{Tup}[\phi S] \to \mathcal{D}_{\leq}(\mathbf{Tup}[\phi T])$ and $g\colon \mathbf{Tup}[\phi U] \to \mathcal{D}_{\leq}(\mathbf{Tup}[\phi V])$ can be composed in parallel if $\phi S \cap \phi U = \phi T \cap \phi V$, in which case

$$f \oplus g\colon \mathbf{Tup}[\phi(S \cup U)] \to \mathcal{D}_{\leq}(\mathbf{Tup}[\phi(T \cup V)])$$

is defined on each $m \in \mathbf{Tup}[\phi(S \cup U)]$ as

$$(f \oplus g)(m)\colon f(m^{\phi S}) \bowtie g(m^{\phi U}) \tag{15}$$

Spelling out the definition of $\bowtie$ for $\mathcal{D}_{\leq}$ (see Section 2), it means that for $n \in \mathbf{Tup}[\phi(T \cup V)]$,

$$(f \oplus g)(m)(n) \coloneqq f(m^{\phi S})(n^{\phi T}) \cdot g(m^{\phi U})(n^{\phi V}) \tag{16}$$

Finally, given two kernels $f\colon \mathbf{Tup}[\phi S] \to \mathcal{D}_{\leq}(\mathbf{Tup}[\phi T])$ and $g\colon \mathbf{Tup}[\phi U] \to \mathcal{D}_{\leq}(\mathbf{Tup}[\phi V])$, $f \sqsubseteq g$ if there exists $W \subseteq_f Var$ and a kernel h such that $g = (f \otimes \eta^{\mathcal{D}}_{\mathbf{Tup}[W]}) \odot h$. We illustrate these definitions with a concrete example.

Example 4.13. Let $z, x, y \in Var$ satisfy $x \preceq y \preceq z$, and ϕ is defined on z, x, y as

$$z \mapsto Z \coloneqq \{0_z, 1_z\} \quad x \mapsto X \coloneqq \{0_x, 1_x\} \quad y \mapsto Y \coloneqq \{0_y, 1_y\}$$

We define two kernels $h_x : \mathbf{Tup}[\{Z\}] \to \mathbf{Tup}[\{Z, X\}]$ and $h_y : \mathbf{Tup}[\{Z\}] \to \mathbf{Tup}[\{Z, Y\}]$ as

$$h_x : 0_z \mapsto {}^1\!/\!_2|0_z 0_x\rangle + {}^1\!/\!_2|0_z 1_x\rangle \qquad h_y : 0_z \mapsto {}^1\!/\!_4|0_z 0_y\rangle + {}^3\!/\!_4|0_z 1_y\rangle$$
$$1_z \mapsto {}^1\!/\!_2|1_z 0_x\rangle + {}^1\!/\!_3|1_z 1_x\rangle \qquad 1_z \mapsto {}^1\!/\!_4|1_z 0_y\rangle + {}^3\!/\!_4|1_z 1_y\rangle$$

Then $h_x \oplus h_y$ is defined, such that for $i_z \in \{0_z, 1_z\}$, $(h_x \oplus h_y)(i_z) = h_x(i_z) \bowtie h_y(i_z)$. For instance,

$$\begin{aligned}
&(h_x \oplus h_y)(0_z)\\
&= h_x(0_z) \bowtie h_y(0_z)\\
&= \left(\frac{1}{2}|0_z 0_x\rangle + \frac{1}{2}|0_z 1_x\rangle\right) \bowtie \left(\frac{1}{4}|0_z 0_y\rangle + \frac{3}{4}|0_z 1_y\rangle\right)\\
&= \frac{1}{8}|0_z 0_x 0_y\rangle + \frac{3}{8}|0_z 0_x 1_y\rangle + \frac{1}{8}|0_z 1_x 0_y\rangle + \frac{3}{8}|0_z 1_x 1_y\rangle
\end{aligned}$$

If we restrict ourselves to proper probability distributions (namely let $\mathcal{T} = \mathcal{D}$), and choose a specific $\phi : Var \to \mathbf{Set}$ such that $\phi s = Val \times \{s\}$ for some fixed set of values Val, then $\mathbf{Tup}[\phi S]$ is isomorphic to $\mathbf{Mem}[S] := \{m : S \to Val\}$. The resulting frame $\langle \mathsf{Ker}_{\le}(\mathcal{K}\ell(\mathcal{D})), \sqsubseteq, \oplus, \odot, \mathsf{Ker}_{\le}(\mathcal{K}\ell(\mathcal{D}))\rangle$ is exactly the probabilistic DIBI frame in Bao et al. [3].

Relational frame. Let $\mathcal{T}$ be the non-empty powerset monad $\mathcal{P}_i$, then apply the same procedure in the previous discussion for subprobabilistic frames and we construct a relational DIBI frame whose kernels are functions $f : \mathbf{Tup}[\phi S] \to \mathcal{P}_i(\mathbf{Tup}[\phi T])$ for some $S \subseteq T$, such that for arbitrary $m \in \mathbf{Tup}[\phi S]$ and $n \in f(m)$, $n^{\phi S} = m$ (namely the projection of n on ϕS is m, see Section 2). In particular, by taking $\phi : (s \in Var) \mapsto Val \times \{s\}$, we recover the relational DIBI frame in Bao et al. [3].

Our categorical framework also enable us to easily derive some interesting properties of the probabilistic and relational frames. For example, in both the probabilistic and relational frames, one kernel can be retrieved from its extensions.

Corollary 4.14. *In both the probabilistic and relational frames, if kernels f and g satisfy $f \sqsubseteq g$, then f can be determined by g: there exist $\mathcal{K}\ell(\mathcal{D})$-morphisms (resp. $\mathcal{K}\ell(\mathcal{P}_i)$-morphisms) h_{pre} and h_{post} such that $f = h_{post} \circ g \circ h_{pre}$.*

While Bao, et al. [3] derived this fact by an extensive computation, this result follows immediately from Proposition 4.5 by observing that $\mathcal{K}\ell(\mathcal{D})$ and $\mathcal{K}\ell(\mathcal{P}_i)$ are indeed Markov categories with extra structure.

Proof of Corollary 4.14. Note that $\mathcal{K}\ell(\mathcal{D})$ and $\mathcal{K}\ell(\mathcal{P}_i)$ are both Markov categories. As for the $\bullet\!-$ structure, in $\mathcal{K}\ell(\mathcal{P}_i)$ we take $\bullet\!-_X : (* \in \mathbf{1}) \mapsto X$ for arbitrary non-empty X; in $\mathcal{K}\ell(\mathcal{P}_i)$, for each finite set $X = \{x_1, \dots, x_m\}$, we take $\bullet\!-_X : (* \in \mathbf{1}) \mapsto {}^1\!/\!_m|x_1\rangle + \cdots + {}^1\!/\!_m|x_m\rangle$. The restriction to non-empty or finite sets is innocuous in our setting. □

Multiset frame. Next, we introduce a new DIBI model we call the multiset model. This model has as states 'input-preserving' functions that output multiset relations. Fix an arbitrary semiring $\mathcal{M} = (M, +, 0, \times, 1)$. The multiset frame will be based on $\mathcal{K}\ell(\mathcal{T}_{\mathcal{M}})$. First, we check that it is indeed a CD category so that we can apply Corollary 4.11. Since $\mathcal{K}\ell(\mathcal{T}_{\mathcal{M}})$ may not be affine for arbitrary semiring $\mathcal{M}$, we cannot apply Lemma 4.12.

Lemma 4.15. *For a semiring $\mathcal{M}$, the Kleisli category $\mathcal{K}\ell(\mathcal{T}_{\mathcal{M}})$ is a CD category.*

Proof. Note that we can still follow the proof of Lemma 4.12 to show that $\mathcal{K}\ell(\mathcal{T})$ is a symmetric monoidal structure. But since $\mathcal{T}$ may not be affine (namely $\mathcal{T}\mathbf{1}$ is not necessarily isomorphic to $\mathbf{1}$), we need to define the CD structure explicitly: given a set X,

- the copier $-\!\!<_X : X \to \mathcal{T}(X \times X)$ maps $x \in X$ to $1|(x, x)\rangle$;
- the discarder $-\!\!\bullet_X : X \to \mathcal{T}(\mathbf{1})$ maps $x \in X$ to $1|*\rangle$.

Then it is straightforward to verify the equations for CD categories. □

Then Corollary 4.11 immediately gives a DIBI frame whose state space consists of kernels in $\mathcal{K}\ell(\mathcal{T}_{\mathcal{M}})$ whose (co)domains obey $\le$. We will explain the details of this DIBI frame below.

The set $\mathsf{Ker}_{\le}(\mathcal{T}_{\mathcal{M}})$ of multiset kernels consists of functions $f : \mathbf{Tup}[\phi S] \to \mathcal{T}_{\mathcal{M}}(\mathbf{Tup}[\phi T])$ for some finite $S \subseteq T$ satisfying that for all $m \in \mathbf{Tup}[\phi S]$ and $n \in supp(f(m))$, $m(\phi s) = n(\phi s)$ for all $\phi s \in \phi S$. Sequential composition of kernels is exactly Kleisli composition. As for parallel composition, given $f : \mathbf{Tup}[\phi S] \to \mathcal{T}_{\mathcal{M}}(\mathbf{Tup}[\phi T])$ and $g : \mathbf{Tup}[\phi U] \to \mathcal{T}_{\mathcal{M}}(\mathbf{Tup}[\phi V])$, $f \oplus g$ is well-defined if $\phi S \cap \phi U = \phi T \cap \phi V$ (equivalently, $S \cap U = T \cap V$ since ϕ is injective), and in this case we define

$$(f \oplus g)(m) = f(m^{\phi S}) \bowtie g(m^{\phi U}) \tag{17}$$
$$(f \oplus g)(m) : n \mapsto f(m^{\phi S})(n^{\phi T}) \times g(m^{\phi U}) g(n^{\phi V}) \tag{18}$$

for all $m \in \mathbf{Tup}[\phi(S \cup U)]$ and $n \in \mathbf{Tup}[\phi(T \cup V)]$. We define $f \sqsubseteq g$ if there exist $W \subseteq_f Var$ and multiset kernel h such that $g = (f \otimes \eta^{\mathcal{T}_{\mathcal{M}}}_{\phi W}) \odot h$. In particular, by taking $\mathcal{M}$ to be $2 = \langle \{0, 1\}, \vee, 0, \wedge, 1\rangle$, the resulting multiset frame is exactly the relational frame. We provide more intuition of multiset kernels and their compositions by an example.

Example 4.16. We consider a scenario of a two machines that produce boxes with treats. One may input one combination of boxes with treats, and they will output multiple combinations of boxes with treats, but all the output boxes keep the original boxes and treats in the input. For instance, one machine receives as input one red box with candies, and outputs one combination of a red box with candies and a green box of chocolates, as well as two combinations of a red box with candies and a green box of cookies. Formally, let $\mathcal{M}$ be the semiring $\mathbb{N}$ of natural numbers with addition and multiplication. Let Var to be $\mathsf{Color} = \{\mathsf{red}, \mathsf{blue}, \mathsf{green}, \mathsf{yellow}\}$ plus some linear ordering $\le$, $\phi(x) = \mathsf{Gift} \times \{x\}$ for all $x \in \mathsf{Color}$ where $\mathsf{Gift} = \{\mathsf{candy}, \mathsf{choco}, \mathsf{cookie}\}$. Then $(\mathcal{K}\ell(\mathcal{T}_{\mathbb{N}}), \le)$-morphisms can equivalently be seen as functions $\mathbf{Mem}[S] \to \mathcal{T}_{\mathbb{N}}(\mathbf{Mem}[T])$, where S, T are finite subsets of Color, $\mathbf{Mem}[S] = \{t\colon S \to \text{Gift}\}$. Consider two kernels

$$f\colon \mathbf{Mem}[\{\text{red}\}] \to \mathcal{T}_{\mathbb{N}}(\mathbf{Mem}[\{\text{red}, \text{green}\}])$$
$$g\colon \mathbf{Mem}[\{\text{red}, \text{blue}\}] \to \mathcal{T}_{\mathbb{N}}(\mathbf{Mem}[\{\text{red}, \text{blue}, \text{yellow}\}])$$

whose components contain, for instance

$$f\colon (\text{red} \mapsto \text{candy}) \mapsto 1|\text{red} \mapsto \text{candy}, \text{green} \mapsto \text{choco}\rangle + 3|\text{red} \mapsto \text{candy}, \text{green} \mapsto \text{cookie}\rangle$$
$$g\colon (\text{red} \mapsto \text{candy}, \text{blue} \mapsto \text{choco}) \mapsto 2|\text{red} \mapsto \text{candy}, \text{blue} \mapsto \text{choco}, \text{yellow} \mapsto \text{cookie}\rangle$$

Then $f \oplus g$ is well-defined, and it is a kernel of type

$$\mathbf{Mem}[\{\text{red}, \text{blue}\}] \to \mathcal{T}_{\mathbb{N}}(\mathbf{Mem}[\{\text{red}, \text{blue}, \text{green}, \text{yellow}\}])$$

For example, its component at $(\text{red} \mapsto \text{candy}, \text{blue} \mapsto \text{choco})$ is

$$2|\text{red} \mapsto \text{candy}, \text{blue} \mapsto \text{choco}, \text{green} \mapsto \text{choco}, \text{yellow} \mapsto \text{cookie}\rangle$$
$$+6|\text{red} \mapsto \text{candy}, \text{blue} \mapsto \text{choco}, \text{green} \mapsto \text{cookie}, \text{yellow} \mapsto \text{cookie}\rangle$$

4.3 Conditional independence in DIBI models

In this section we turn to conditional independence in categorical models based on the categorical frames defined in 4.1. We define two notions of CI for these categorical models — one via DIBI formula, one via string diagrams — and prove their equivalence.

We first define our logic $\mathrm{DIBI}_{\{\wedge,*,\fatsemi\}}$. Let *Var* be a set of variables linearly ordered by $\leq$. $\mathrm{DIBI}_{\{\wedge,*,\fatsemi\}}$ is the conjunctive fragment of DIBI, and has atomic propositions of the form $(A \triangleright [B])$. Informally, the formula $(A \triangleright [B])$ describes a kernel whose domain is exactly A and codomain contains B, thus B may depend on A in this kernel.

Definition 4.17. The syntax of $\mathrm{DIBI}_{\{\wedge,*,\fatsemi\}}$ formulas are inductively defined as follows:

$$P, Q ::= p \in \mathcal{AP} \mid \top \mid I \mid P \wedge Q \mid P * Q \mid P \fatsemi Q$$

where the set $\mathcal{AP}$ of atomic propositions is $\{(A \triangleright [B]) \mid A, B \text{ are finite subsets of } Var\}$.

Every categorical DIBI frame $\mathrm{Ker}(\mathbf{C})$ in Theorem 4.8 together with an injective function $\phi\colon Var \to \mathbf{Ob}(\mathbf{C})$ determines a model for $\mathrm{DIBI}_{\{\wedge,*,\fatsemi\}}$, using the construction of $\mathrm{Ker}_{\leq}(\mathbf{C})$ in Corollary 4.11.

Definition 4.18 (Categorical DIBI model)**.** Every injective $\phi\colon Var \to \mathbf{Ob}(\mathbf{C})$ determines a $\mathrm{DIBI}_{\{\wedge,*,\fatsemi\}}$ model $\langle \mathrm{Ker}_{\leq}(\mathbf{C}), \sqsubseteq, \oplus, \odot, \mathrm{Ker}_{\leq}(\mathbf{C}), \mathcal{V}\rangle$, where

- $\langle \mathrm{Ker}_{\leq}(\mathbf{C}), \sqsubseteq, \oplus, \odot, \mathrm{Ker}_{\leq}(\mathbf{C})\rangle$ is the DIBI frame in Corollary 4.11,
- $\mathcal{V}\colon \mathcal{AP} \to \mathcal{P}(\mathrm{Ker}_{\leq}(\mathbf{C}))$ defines $\mathcal{V}((S \triangleright [T]))$ to be

$$\{f \in \mathrm{Ker}_{\leq}(\mathbf{C}) \mid \exists f' \sqsubseteq f \text{ s.t. } dom(f') = S, cod(f') \supseteq T\}$$

Definition 4.19 ($\mathrm{DIBI}_{\{\wedge,*,\fatsemi\}}$ satisfaction)**.** Satisfaction $\vDash_{\mathcal{V}}$ of $\mathrm{DIBI}_{\{\wedge,*,\fatsemi\}}$ formulas at a kernel f is inductively defined by the clauses in Figure 4.

$f \vDash_{\mathcal{V}} (A \triangleright [B])$	iff	$f \in \mathcal{V}((A \triangleright [B]))$
$f \vDash_{\mathcal{V}} \top$	iff	always
$f \vDash_{\mathcal{V}} I$	iff	always[2]
$f \vDash_{\mathcal{V}} P \wedge Q$	iff	$f \vDash_{\mathcal{V}} P$ and $f \vDash_{\mathcal{V}} Q$
$f \vDash_{\mathcal{V}} P * Q$	iff	$\exists g, h$ s.t. $g \oplus h \sqsubseteq f, g \vDash_{\mathcal{V}} P, h \vDash_{\mathcal{V}} Q$
$f \vDash_{\mathcal{V}} P \fatsemi Q$	iff	$\exists g, h$ s.t. $g \odot h = f, g \vDash_{\mathcal{V}} P, h \vDash_{\mathcal{V}} Q$

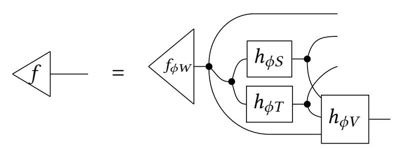
Figure 4. Satisfaction in $\mathrm{DIBI}_{\{\wedge,*,\fatsemi\}}$

Bao et al. [3] show that certain $\mathrm{DIBI}_{\{\wedge,*,\fatsemi\}}$ formulas characterise both conditional independence in probabilistic DIBI models and join dependency in relational DIBI models. This motivates defining a notion of logical CI (thus the subscript 'L' in $\perp\!\!\!\perp_L$ below) in terms of satisfaction of such formulas.

Definition 4.20 (DIBI conditional independence)**.** Let S, T, W, V be finite subsets of *Var*, and $f\colon I \to \phi[S \cup T \cup W \cup V]$ be a kernel. f satisfies that S and T are *DIBI conditionally independent* given W, denoted as $S \perp\!\!\!\perp_L T \mid W$, if

$$f \vDash_{\mathcal{V}} (\varnothing \triangleright [W]) \fatsemi ((W \triangleright [S]) * (W \triangleright [T]))$$

Next we adapt the notion of process conditional independence (Definition 3.1) to the categorical DIBI models. Under the assumption in Definition 4.20, f satisfies that ϕS and ϕT are process conditionally independent given ϕW if $S \cap T \subseteq W$, there exist $f_{\phi W}$, $h_{\phi S}$ and $h_{\phi T}$ with $cod(f_{\phi W}) = \phi W$, $dom(h_{\phi S}) = dom(h_{\phi T}) = \phi W$, $cod(h_{\phi S}) \supseteq \phi[S \setminus W]$, $cod(h_{\phi T}) \supseteq \phi[T \setminus W]$ satisfying



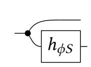
(19)

For readability, when ϕ is clear from the context, we write $S \perp\!\!\!\perp_P T \mid W$ for $\phi S \perp\!\!\!\perp_P \phi T \mid \phi W$. These two notions of CI — the string diagrammatic one and the logical one — for our categorical models coincide:

Theorem 4.21. *For a kernel* $f\colon I \to \phi(S \cup T \cup W \cup V)$,

- f *satisfies* $S \perp\!\!\!\perp_P T \mid W$ *implies* f *satisfies* $S \perp\!\!\!\perp_L T \mid W$*;*
- *if in addition* $\mathbf{C}$ *is a Markov category with conditionals, then* f *satisfies* $S \perp\!\!\!\perp_P T \mid W$ *if and only if* f *satisfies* $S \perp\!\!\!\perp_L T \mid W$.

Proof sketch. For the direction from $\perp\!\!\!\perp_P$ to $\perp\!\!\!\perp_L$, the key observation is that the right-hand-side of (19) is $f_{\phi W} \odot (f_{\phi S} \oplus f_{\phi T}) \odot f_{\phi V}$, where for example $f_{\phi S} = (id_{\phi W} \otimes h_{\phi S}) \circ \blacktriangleleft_{\phi W}$, or graphically,



Then $f \vDash_{\mathcal{V}} (\varnothing \triangleright [W]) \fatsemi ((W \triangleright [S]) * (W \triangleright [T]))$ follows immediately from the definition of $\mathrm{DIBI}_{\{\wedge,*,\fatsemi\}}$ satisfaction.

As for the other direction, we clarify why the extra conditional of Markov categories with conditioanls is necessary. $f \vDash_{\mathcal{V}} (\varnothing \triangleright [W]) \fatsemi ((W \triangleright [S]) * (W \triangleright [T]))$ means that there exist kernels f_0, h_S, h_T such that $f_0 \vDash_{\mathcal{V}} (\varnothing \triangleright [W])$, $h_S \vDash_{\mathcal{V}} (W \triangleright [S])$, $h_T \vDash_{\mathcal{V}} (W \triangleright [T])$ and $f_0 \odot (h_S \oplus h_T) \sqsubseteq f$. Then by Definition 4.19 $cod(f_0) \supseteq W$. But this does not imply $S \perp\!\!\!\perp_P T \mid W$ because there is no guarantee that $cod(f_0)$ is exactly W. If $\mathbf{C}$ is a Markov category with conditionals, then this induce a f_W whose codomain is exactly W. □

As an application, we recover the characterisation theorems for CI in probabilistic frames and for join dependency in relational frames in [3] by instantiating to the probabilistic and (multiset) relations frames.

Corollary 4.22. *Given arbitrary finite sets $S, T, W, U \subseteq Var$ satisfying $S \cap T \subseteq W$ and a distribution $\mu \in \mathcal{D}(\mathbf{Tup}[S \cup T \cup W \cup U])$, μ satisfies $S \perp\!\!\!\perp_{pr} T \mid W$ if and only if $\mu \vDash_{\mathcal{V}} (\varnothing \triangleright [W]) \fatsemi ((W \triangleright [S]) * (W \triangleright [T]))$.*

Corollary 4.23. *Given arbitrary finite sets S, T and a $\mathcal{M}$-multiset relation $R \in \mathcal{T}_{\mathcal{M}}(\mathbf{Tup}[S \cup T])$, R satisfies join dependency $S \bowtie T$ if and only if $R \vDash_{\mathcal{V}} (\varnothing \triangleright [S \cap T]) \fatsemi ((S \cap T \triangleright [S]) * (S \cap T \triangleright [T]))$.*

Proof sketch of Corollary 4.22. Note that $\mathcal{K}\ell(\mathcal{D})$ is a Markov category with conditionals. Since $S \perp\!\!\!\perp_{pr} T \mid W$ and kernels in probability models are special cases of $S \perp\!\!\!\perp_P T \mid W$ and $(\mathbf{C}, \preceq)$-kernels in the setting of $\mathcal{K}\ell(\mathcal{D})$, respectively, the equivalence follows immediately from Theorem 4.21. □

5 Related work

Logics of independence and conditional independence. There are several logical treatments that formalize and reason about conditional independence. Pearl and Paz [18] showed that several natural notions of conditional independence satisfied the *graphoid* axioms, and showed that relations and probability distributions were two concrete models. More recent approaches include *Independence-friendly logic* (IF) [12] and dependence logic [21]. Both extend classical first-order-logic by introducing new quantifiers to express that a certain variable does not or only relies on a set of variable. Surprisingly, compositional semantics of IF was not known until Hodges' team semantics [13, 14], and the natural propositional logic associated with Hodges' construction is bunched implication (BI) [1], whose categorical models are single categories with two closed structure: one cartesian and one symmetric monoidal. Their connection with DIBI from a categorical semantics point of view is worth exploring.

CI in string diagrams. To the best of our knowledge, the first string diagrammatic treatment of CI is from [7] in the setting of (quantum) Bayesian graphical calculus. A diagrammatic definition of CI like (3) is first proposed by Cho and Jacobs [5] in the context of Markov categories with conditional. Fritz [10] generalise this definition in absence of conditionals. Notably, all these works are not restricted to CI in 'distributions' $I \to X$, and extensions to arbitrary morphisms $U \to X$ is straightforward. We believe this is also the case for our definitions.

There are other non-graphical categorical treatment of CI [9, 11, 20]. A discussion about the relation between [9, 11] and the diagrammatic definitions of CI can be found in [10]. Section 3.1 is a first step towards relating [20] and ST-CI.

6 Discussion and future directions

Notions of conditional independence have been studied from a categorical and logical perspectives in the literature. In this paper, we bring these two perspectives closer to each other by providing general categorical models for a logic of (in)dependence—DIBI. Our categorical models make use of string diagrammatic reasoning and are general enough to, on the one hand, recover existing models of DIBI (probabilistic and relational) and, on the other hand, provide new models, as for instance the multiset model we presented in Section 4.

We see several possible directions for future work. The first direction is exploring how to use DIBI for reasoning about programs. DIBI was originally designed to serve as a logic of assertions for a program logic, analogously to the role of the logic of bunched implications in separation logic. In [3] a first attempt was made at designing a program logic for the probabilistic model but the proof rules were intricate and hard to use in even simple programs. We believe the string diagrammatic perspective we present in this paper may offer an alternative, simpler way to design a program logic for probabilistic programs. Furthermore, our categorical model of DIBI gives rise to many new concrete models and it could be interesting to investigate whether DIBI could be used to reason about other kinds of programs. Since the prime example of DIBI models are based on Kleisli arrows for a commutative monad, a natural question is whether these models be useful when reasoning about programs with different kinds of monadic effects, e.g., non-determinism.

The second direction is extending the bridge between string diagrams and bunched logics. The logic BI [17] initiated a wave of interest in bunched logics, with many variants proposed over the last twenty years. More recently, there has been interest in a more uniform treatment of such logics, by defining the syntax and semantics of a broad class of bunched logics so that their metatheory could be studied via generic tools [3]. DIBI is one example of a bunched logic that can be developed in this way. In one direction, it would be interesting to see if string diagrams can also provide a categorical semantics for other bunched logics. In the other direction, we wonder what kinds of string diagrams are needed to characterize the models of standard bunched logics, like BI.

References

[1] Samson Abramsky and Jouko Väänänen. 2009. From if to bi. *Synthese* 167, 2 (2009), 207–230.

[2] Alfred V Aho, Catriel Beeri, and Jeffrey D Ullman. 1979. The theory of joins in relational databases. *ACM Transactions on Database Systems (TODS)* 4, 3 (1979), 297–314.

[3] Jialu Bao, Simon Docherty, Justin Hsu, and Alexandra Silva. 2020. A Logic to Reason about Dependence and Independence. *ArXiv* abs/2008.09231 (2020).
[4] Gilles Barthe, Justin Hsu, and Kevin Liao. 2019. A probabilistic separation logic. *Proceedings of the ACM on Programming Languages* 4, POPL (2019), 1–30.
[5] Kenta Cho and Bart Jacobs. 2019. Disintegration and Bayesian inversion via string diagrams. *Mathematical Structures in Computer Science* 29, 7 (2019), 938–971.
[6] Florence Clerc, Vincent Danos, Fredrik Dahlqvist, and Ilias Garnier. 2017. Pointless learning. In *International Conference on Foundations of Software Science and Computation Structures*. Springer, 355–369.
[7] Bob Coecke and Robert W Spekkens. 2012. Picturing classical and quantum Bayesian inference. *Synthese* 186, 3 (2012), 651–696.
[8] A Philip Dawid. 1979. Conditional independence in statistical theory. *Journal of the Royal Statistical Society: Series B (Methodological)* 41, 1 (1979), 1–15.
[9] Uwe Franz. 2002. What is stochastic independence? In *Non-commutativity, infinite-dimensionality and probability at the crossroads*. World Scientific, 254–274.
[10] Tobias Fritz. 2020. A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics* 370 (Aug 2020), 107239. https://doi.org/10.1016/j.aim.2020.107239
[11] Malte Gerhold, Stephanie Lachs, and Michael Schürmann. 2016. Categorial L\'evy Processes. *arXiv preprint arXiv:1612.05139* (2016).
[12] Jaakko Hintikka. 1998. Informational independence as a semantical phenomenon. In *Paradigms for Language Theory and Other Essays*. Springer, 52–70.
[13] Wilfrid Hodges. 1997. Compositional semantics for a language of imperfect information. *Logic Journal of the IGPL* 5, 4 (1997), 539–563.
[14] Wilfrid Hodges. 1997. Some strange quantifiers. In *Structures in logic and computer science*. Springer, 51–65.
[15] Bart Jacobs. 1994. Semantics of weakening and contraction. *Annals of pure and applied logic* 69, 1 (1994), 73–106.
[16] Bart Jacobs and Fabio Zanasi. 2017. A formal semantics of influence in Bayesian reasoning. (2017).
[17] Peter W O'Hearn and David J Pym. 1999. The logic of bunched implications. *Bulletin of Symbolic Logic* 5, 2 (1999), 215–244.
[18] Judea Pearl. 2009. *Causality*. Cambridge university press.
[19] Peter Selinger. 2010. A survey of graphical languages for monoidal categories. In *New structures for physics*. Springer, 289–355.
[20] Alex Simpson. 2018. Category-theoretic Structure for Independence and Conditional Independence. *Electronic Notes in Theoretical Computer Science* 336 (2018), 281–297. https://doi.org/10.1016/j.entcs.2018.03.028 The Thirty-third Conference on the Mathematical Foundations of Programming Semantics (MFPS XXXIII).
[21] Jouko Väänänen. 2007. *Dependence logic: A new approach to independence friendly logic*. Vol. 70. Cambridge University Press.

A Omitted proofs

A.1 Missing proofs for Theorem 4.8

This appendix includes the proof of Theorem 4.8 and some lemmas for the proof (stated after the main proof). Since (co)domains of the kernels are tensor products of distinct objects, we may regard $dom(f)$ as the set $\{X_1, \dots, X_n\}$ if f is of the form $X_1 \otimes \cdots \otimes X_n \to Y$. We may also use set theoretical relations such as $\subseteq$, for example, in $dom(f) \subseteq cod(g)$.

The equations between kernels below will be modolu reordering of domain and codomain wires via χ. This will not be problematic because the (co)domains of kernels are tensors of distinct objects. We use F, G, H... to denote such sloppy variation of kernels. One can retrieve the proof for kernel f, g, h... by spelling out those χ explicitly in the proof.

Proof of Theorem 4.8. We verify the frame conditions in Figure 3 for the categorical frames in Section 4.1.

1. $\oplus$ Down-closed: Suppose $F \oplus G$ is defined, $F' \sqsubseteq F$, $G' \sqsubseteq G$. We first show that $F' \oplus G'$ is defined, namely $dom(F') \cap dom(G') = cod(F') \cap cod(G')$. Note that by Lemma A.2, the condition for $F \oplus G$ is equivalent to that $dom(F) \cap (cod(G) \setminus dom(G)) = dom(G) \cap (cod(F) \setminus dom(F)) = (cod(F) \setminus dom(F)) \cap (cod(G) \setminus dom(G)) = \emptyset$.
   a. $F' \sqsubseteq F$ and $G' \sqsubseteq G$ imply that $dom(F') \subseteq dom(F)$ and $cod(G') \setminus dom(G) \subseteq cod(G) \setminus dom(G)$ (by Lemma A.4). Then $dom(F') \cap (cod(G') \setminus dom(G)) \subseteq dom(F) \cap (cod(G) \setminus dom(G)) = \emptyset$, which means $dom(F') \cap (cod(G') \setminus dom(G')) = \emptyset$.
   b. Using a similar argument as that For (a), we have $dom(G') \cap (cod(F') \setminus dom(F')) = \emptyset$.
   c. By Lemma A.4, $(cod(F') \setminus dom(F')) \cap (cod(G') \setminus dom(G')) \subseteq (cod(F) \setminus dom(F)) \cap (cod(G) \setminus dom(G)) = \emptyset$.

   Next, we show $F' \oplus G' \sqsubseteq F \oplus G$. Suppose $F = (F' \oplus id_S) \odot V$, $G = (G' \oplus id_T) \odot W$. Then

$$\begin{aligned} F \oplus G &= ((F' \oplus id_S) \odot V) \oplus ((G' \oplus id_T) \odot W) \\ &= (F' \oplus id_S) \oplus (G' \oplus id_T) \odot (V \oplus W) \\ &= (F' \oplus G') \oplus (id_S \oplus id_T)) \odot (V \oplus W) \\ &= ((F' \oplus G') \oplus (id_{(S \cup T)}) \odot (V \oplus W) \\ &\sqsupseteq F' \oplus G' \end{aligned}$$

2. $\odot$ Up-closed: Suppose $F \odot G \sqsubseteq H'$, so there exists object X and kernel K such that $H' = ((F \odot G) \oplus (id_X) \odot K$.

$$\begin{aligned} H' &= ((F \odot G) \oplus id_X) \odot K \\ &= ((F \oplus id_X) \odot (G \oplus id_X)) \odot K \\ &= (F \oplus id_X) \odot (G \oplus id_X) \odot K) \end{aligned}$$

   Let $F' = F \oplus id_X$, and $G' = (G \oplus id_X) \odot K$, then $H = F' \odot G'$.

3. $\oplus$ Commutativity: It immediately follows from the definition of $\oplus$ that the order of the two components for $\oplus$ does not matter.
4. $\oplus$ Associativity: Suppose $F \oplus G$ and $(F \oplus G) \oplus H$ are both defined. We first show that $G \oplus H$ and $F \oplus (G \oplus H)$ are also defined. By Corollary A.3, $F \oplus G$ and $(F \oplus G) \oplus H$ imply $cod(F) \cap (cod(G) \setminus dom(G)) = cod(G) \cap (cod(F) \setminus dom(F)) = \emptyset$ and $cod(F \oplus G) \cap (cod(H) \setminus dom(H)) = cod(H) \cap (cod(F \oplus G) \setminus dom(F \oplus G)) = \emptyset$.
   a. $cod(G) \subseteq cod(F) \cup cod(G) = cod(F \oplus G)$ implies that $cod(G) \cap (cod(H) \setminus dom(H)) \subseteq cod(F \oplus G) \cap (cod(H) \setminus dom(H)) = \emptyset$. Using Lemma A.5, we know that $cod(H) \cap (cod(G) \setminus dom(G)) \subseteq cod(H) \cap (cod(F \oplus G) \setminus dom(F \oplus G)) = \emptyset$. Then, $cod(G) \cap (cod(H) \setminus dom(H)) = cod(H) \cap (cod(G) \setminus dom(G)) = \emptyset$ implies that $G \oplus H$ is defined, via Corollary A.3.
   b. By Lemma A.6, $cod(F) \cap (cod(G \oplus H) \setminus dom(G \oplus H)) = cod(F) \cap ((cod(G) \setminus dom(G)) \cup (cod(H) \setminus dom(H))) = (cod(F) \cap (cod(G) \setminus dom(G))) \cup (cod(F) \cap (cod(H) \setminus dom(H))) = cod(F) \cap (cod(H) \setminus dom(H)) \subseteq cod(F \oplus G) \cap (cod(H) \setminus dom(H)) = \emptyset$. Also, $cod(G \oplus H) \cap (cod(F) \setminus dom(F)) = (cod(G) \cup cod(H)) \cap (cod(F) \setminus dom(F)) = (cod(G) \cap (cod(F) \setminus dom(F))) \cup (cod(H) \cap (cod(F) \setminus dom(F)))$. Note that $cod(G) \cap (cod(F) \setminus dom(F)) = \emptyset$, $cod(H) \cap (cod(F) \setminus dom(F)) \subseteq cod(H) \cap (cod(F \oplus G) \setminus dom(F \oplus G)) = \emptyset$ (Lemma A.5), so $cod(G \oplus H) \cap (cod(F) \setminus dom(F)) = \emptyset$. Then it follows that $F \oplus (G \oplus H)$ is defined.

   The equivalence of $(F \oplus G) \oplus H$ and $F \oplus (G \oplus H)$ is immediate from their diagrammatic presentation.
5. $\oplus$ Unit existence: for arbitrary kernel F, $F \oplus id_I = F$.
6. $\oplus$ Unit Coherence: already shown in the main text.
7. $\odot$ Associativity: immediate from associativity of composition $\circ$ in categories.
8. $\odot$ Unit Existence$_\mathrm{L}$: For arbitrary $F\colon X \to Y$, id_X is also a kernel, and $id_X \odot F = F$.
9. $\odot$ Unit Existence$_\mathrm{R}$: For arbitrary $F \in \mathsf{Ker}(\mathcal{T})$ of type $X \to Y$, id_Y is also in $\mathsf{Ker}(\mathcal{T})$, and $F \odot id_Y = F$.
10. $\odot$ Coherence$_\mathrm{R}$: For arbitrary F, G, E, suppose $F = G \odot E$, then it is by definition of $\sqsubseteq$ that $G \sqsubseteq F$.
11. Unit closure: Since the unit set is $\mathsf{Ker}(\mathcal{T})$ itself, this condition holds trivially.
12. Reverse exchange: For arbitrary $F_1, F_2, G_1, G_2 \in \mathsf{Ker}(\mathcal{T})$, suppose $(F_1 \odot F_2) \oplus (G_1 \odot G_2)$ is defined, we show that $(F_1 \oplus G_1) \odot (F_2 \oplus G_2)$ is also defined, and indeed equivalent to $(F_1 \odot F_2) \oplus (G_1 \odot G_2)$. From the assumption, we know that $cod(F_1) = dom(F_2)$, $cod(G_1) = dom(G_2)$, $dom(F_1) \cap dom(G_1) = cod(F_2) \cap cod(G_2)$.
   - $F_1 \oplus G_1$ is defined: $dom(F) \cap dom(G) \subseteq cod(F) \cap cod(G)$ is always true for arbitrary $F, G \in \mathsf{Ker}(\mathcal{T})$. For the other inclusion, note that $cod(F_1) \cap cod(G_1) = dom(F_2) \cap dom(G_2) \subseteq cod(F_2) \cap cod(G_2) = dom(F_1) \cap dom(G_1)$. Together we have $dom(F_1) \cap dom(G_1) = cod(F_1) \cap cod(G_1)$.
   - $F_2 \oplus G_2$ is defined: Similar to the previous argument, we have $dom(F_2) \cap dom(G_2) = cod(F_1) \cap cod(G_1) \supseteq dom(F_1) \cap dom(G_1) = cod(F_2) \cap cod(G_2)$.
   - $(F_1 \oplus G_1) \odot (F_2 \oplus G_2)$ is defined: $cod(F_1 \oplus G_1) = cod(F_1) \cup cod(G_1) = dom(F_2) \cup dom(G_2) = dom(F_2 \oplus G_2)$.

   Then the equivalence of $(F_1 \odot F_2) \oplus (G_1 \odot G_2)$ and $(F_1 \oplus G_1) \odot (F_2 \oplus G_2)$ follows immediately from their diagrammatic presentations.

□

The following lemma gives a handy equivalent condition for $f \oplus g$ being well-defined. We abstract away the concrete (co)domains and use arbitrary sets.

Lemma A.1. *Suppose* $X_0 \subseteq X_1$, $Y_0 \subseteq Y_1$, *then* $X_0 \cap Y_0 = X_1 \cap Y_1$ *if and only if* $X_0 \cap (Y_1 \setminus Y_0) = Y_0 \cap (X_1 \setminus X_0) = (X_1 \setminus X_0) \cap (Y_1 \setminus Y_0) = \emptyset$.

Proof. Since $X_0 \subseteq X_1$, $Y_0 \subseteq Y_1$, we know that

$$\begin{aligned} X_1 \cap Y_1 &= (X_0 \cup (X_1 \setminus X_0)) \cap (Y_0 \cup (Y_1 \setminus Y_0)) \\ &= (X_0 \cap Y_0) \cup (X_0 \cap (Y_1 \setminus Y_0)) \\ &\quad \cup (Y_0 \cap (X_1 \setminus X_0)) \cup ((X_1 \setminus X_0) \cap (Y_1 \setminus Y_0)) \end{aligned}$$

Then, for the 'if' direction, the three sets being empty immediately implies that $X_1 \cap Y_1 = X_0 \cap Y_0$.

For the 'only if' direction, suppose $X_1 \cap Y_1 = X_0 \cap Y_0$, then

$$(X_0 \cap Y_0) \cup (X_0 \cap (Y_1 \setminus Y_0)) \cup (Y_0 \cap (X_1 \setminus X_0)) \cup ((X_1 \setminus X_0) \cap (Y_1 \setminus Y_0)) = X_0 \cap Y_0 \tag{20}$$

Since $(Y_1 \setminus Y_0) \cap Y_0 = \emptyset$, $X_0 \cap (Y_1 \setminus Y_0)$ is disjoint with $X_0 \cap Y_0$, and (20) implies $X_0 \cap (Y_1 \setminus Y_0) = \emptyset$. Similarly we have $Y_0 \cap (X_1 \setminus X_0) = \emptyset$. Also, both $X_1 \setminus X_0$ and $Y_1 \setminus Y_0$ are disjoint with $X_0 \cap Y_0$, so $((X_1 \setminus X_0) \cap (Y_1 \setminus Y_0)) \cap (X_0 \cap Y_0) = \emptyset$, which together with (20) imply that $(X_1 \setminus X_0) \cap (Y_1 \setminus Y_0) = \emptyset$. □

Lemma A.2. *For arbitrary* $f, g \in \mathsf{Ker}(\mathcal{T})$, $f \oplus g$ *is defined if and only if* $dom(f) \cap (cod(g) \setminus dom(g)) = dom(g) \cap (cod(f) \setminus dom(f)) = (cod(f) \setminus dom(f)) \cap (cod(g) \setminus dom(g)) = \emptyset$.

Proof. Let X_0 and X_1 be $dom(f)$ and $cod(f)$, Y_0 and Y_1 be $dom(g)$ and $cod(g)$, and then apply Lemma A.1. □

Corollary A.3. *For arbitrary* $f, g \in \mathsf{Ker}(\mathcal{T})$, $f \oplus g$ *is defined if and only if* $cod(f) \cap (cod(g) \setminus dom(g)) = cod(g) \cap (cod(f) \setminus dom(f)) = \emptyset$.

Proof. It suffices to show that this condition is equivalent to that in Lemma A.5.

Starting from the condition here, since $dom(f) \subseteq cod(f)$ and $dom(g) \subseteq cod(g)$, it follows immediately that $dom(f) \cap (cod(g) \setminus dom(g)) = dom(g) \cap (cod(f) \setminus dom(f)) = \emptyset$. Also, $cod(f) \setminus dom(f) \subseteq cod(f)$, so $(cod(f) \setminus dom(f)) \cap (cod(g) \setminus dom(g)) \subseteq cod(f) \cap (cod(g) \setminus dom(g)) = \emptyset$.

For the other direction, assume the conditions in Lemma A.2, $cod(f) \cap (cod(g) \setminus dom(g)) = (dom(f) \cup (cod(f) \setminus dom(f))) \cap (cod(g) \setminus dom(g)) = (dom(f) \cap (cod(g) \setminus dom(g))) \cup ((cod(f) \setminus$

$dom(f)) \cap (cod(g) \setminus dom(g))) = \emptyset$. Similarly we have $cod(g) \cap (cod(f) \setminus dom(f)) = \emptyset$. □

Lemma A.4. *Suppose $f, f' \in \mathrm{Ker}(\mathcal{T})$ satisfy $f' \sqsubseteq f$, then $cod(f') \setminus dom(f') \subseteq cod(f) \setminus dom(f)$.*

Proof. By definition, $f' \sqsubseteq f$ means that there exist object X and morphism $h \in \mathrm{Ker}(\mathcal{T})$ such that $f = \sigma_1; (f' \otimes id_X); \sigma_2; h$, where σ_1 and σ_2 both consists of suitably many swapping morphisms $\times$. So $dom(f) = dom(f') \cup X$, $cod(f) = cod(h) = dom(h) \cup (cod(h) \setminus dom(h)) = (cod(f') \cup X) \cup (cod(h) \setminus dom(h))$. Then $cod(f) \setminus dom(f) = (cod(f') \cup X \cup (cod(h) \setminus dom(h))) \setminus (dom(f') \cup X) = (cod(f') \setminus dom(f')) \cup (cod(h) \setminus dom(h))$, which implies $cod(f') \setminus dom(f') \subseteq cod(f) \setminus dom(f)$. □

Lemma A.5. *Suppose $f, g \in \mathrm{Ker}(\mathcal{T})$ and $f \oplus g$ is defined, then $cod(f) \setminus dom(f) \subseteq cod(f \oplus g) \setminus dom(f \oplus g)$.*

Proof. By the definition of $f \oplus g$ being defined, $cod(f \oplus g)$ can be expressed as the union of three disjoint sets $dom(f \oplus g) \cup (cod(f) \setminus dom(f)) \cup (cod(g) \setminus dom(g))$, and $cod(f \oplus g) \setminus dom(f \oplus g) = (cod(f) \setminus dom(f)) \cup (cod(g) \setminus dom(g)) \supseteq (cod(f) \setminus dom(f))$. □

Lemma A.6. *Suppose $f, g \in \mathrm{Ker}(\mathcal{T})$ and $f \oplus g$ is defined, then $cod(f \oplus g) \setminus dom(f \oplus g) = (cod(f) \setminus dom(f)) \cup (cod(g) \setminus dom(g))$.*

Proof. Since $dom(f \oplus g) = dom(f) \cup dom(g)$ and $cod(f \oplus g) = cod(f) \cup cod(g)$,

$$\begin{aligned}
&cod(f \oplus g) \setminus dom(f \oplus g)\\
=&(cod(f) \cup cod(g)) \setminus (dom(f) \cup dom(g))\\
=&(cod(f) \setminus (dom(f) \cup dom(g))) \cup (cod(g) \setminus (dom(f) \cup dom(g)))\\
=&((cod(f) \setminus dom(f)) \setminus dom(g)) \cup ((cod(g) \setminus dom(g)) \cup dom(f))\\
=&(cod(f) \setminus dom(f)) \cup (cod(g) \setminus dom(g)) \qquad (21)
\end{aligned}$$

where (21) is because $(cod(f) \setminus dom(f)) \cap dom(g) = (cod(g) \setminus dom(g)) \cap dom(f) = \emptyset$ (Lemma A.2). □

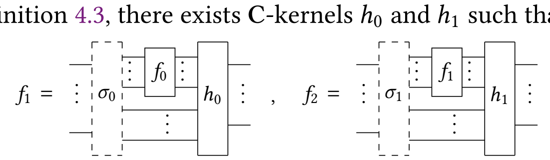
A.2 Other missing proofs

Proposition 3.4. For the 'only if' direction, that f satisfies $X \perp\!\!\!\perp_P Y \mid Z$ means that f can be decomposed as in (4). Since $-\!\bullet$ is natural, discarding the output wire U on both sides of (4) returns exactly (5).

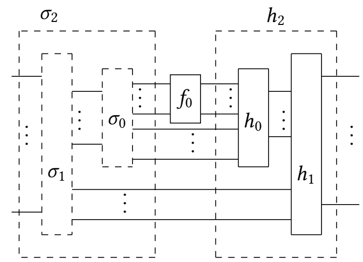
For the 'if' direction, suppose (5) holds, then $\mathbf{C}$ having conditionals implies that there exists h_U such that (4) holds. In particular, $U' = U$, and X', Y' are both empty. □

Proposition 4.4. $\sqsubseteq$ is clearly reflexive: for arbitrary $\mathbf{C}$-kernel, $f = f; id$.

For transitivity, suppose $f_0 \sqsubseteq f_1$ and $f_1 \sqsubseteq f_2$, then by Definition 4.3, there exists $\mathbf{C}$-kernels h_0 and h_1 such that:



This implies that f_2 can be decomposed as:



(22)

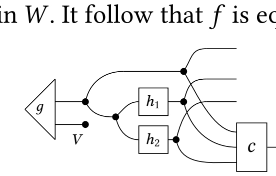
Then σ_2 and h_2 in (22) witness that $f_0 \sqsubseteq f_2$. □

Proof of Lemma 4.12. This follows immediately from (the proof of) Corollary 3.2 from [10], where we change the setting from Markov categories to CD categories. In particular, the monoidal structure $\langle \boxplus, J \rangle$ on $\mathcal{K}\ell(\mathcal{T})$ is defined as follows: on objects, $A \boxplus B := A \otimes B$; on morphisms, given $f: A \to \mathcal{T}B$ and $g: C \to \mathcal{T}D$, $f \boxplus g$ is defined as the composition $A \otimes C \xrightarrow{f \otimes g} \mathcal{T}(B) \otimes \mathcal{T}(D) \xrightarrow{\mathrm{dst}_{B,D}} \mathcal{T}(B \otimes D)$, where dst is the strength of $\mathcal{T}$; the tensor unit J is simply I. So with a bit abuse of notation we may write the symmetric monoidal structure on $\mathcal{K}\ell(\mathcal{T})$ as $\langle \otimes, I \rangle$. The CD structure of $\mathcal{K}\ell(\mathcal{T})$ is inheritated from that of $\mathbf{C}$. In particular, that the discarder being unit of copier follows from that $\mathcal{T}$ is affine. □

Proof of Theorem 4.21. Since (co)domains of the kernels are tensor products of distinct objects, we may regard $dom(f)$ as the set $\{X_1, \ldots, X_n\}$ if f is of the form $X_1 \otimes \cdots \otimes X_n \to Y$.

We assume that f satisfies $S \perp\!\!\!\perp_P T \mid W$ and $S \cap T = W$. By Definition 3.1, there exist $h, f_W, h_Y, h_Z \in \mathrm{Ker}(\mathcal{T})$ such that $f = f_W \odot h$ and $h \sqsupseteq h_S \oplus h_T$, with $cod(f_W) = W$, $dom(h_S) = dom(h_T) = W$, $cod(h_S) \supseteq S$, $cod(h_T) \supseteq T$. First, we immediately have $f_W \vDash_{\mathcal{V}} (\emptyset \triangleright [W])$ $h_S \vDash_{\mathcal{V}} (W \triangleright [S])$ and $h_T \vDash (W \triangleright [T])$. Next, $h \sqsupseteq h_1 \oplus h_2$ implies that $h \vDash_{\mathcal{V}} (W \triangleright [S]) * (W \triangleright [T])$. Then, the above statements together with $f = f_W \odot h$ imply that $f \vDash_{\mathcal{V}} (\emptyset \triangleright [W]) \fatsemi (S \triangleright [T]) * (S \triangleright [T])$.

Now suppose $\mathbf{C}$ is a Markov category with conditional, and we show that $\perp\!\!\!\perp_L$ implies $\perp\!\!\!\perp_P$. Suppose f satisfies $S \perp\!\!\!\perp_L T \mid W$, namely $f \vDash_{\mathcal{V}} (\emptyset \triangleright [W]) \fatsemi (W \triangleright [S]) * (W \triangleright [T])$. By the definition of the semantics, there exist g, h such that $f = g \odot h$, $g \vDash_{\mathcal{V}} (\emptyset \triangleright [W])$, $h \vDash_{\mathcal{V}} (W \triangleright [S]) * (W \triangleright [T])$. So $dom(g) = \mathbf{1}$, $cod(g) \supseteq W$. $h \vDash_{\mathcal{V}} (W \triangleright [S]) * (W \triangleright [T])$ implies that there exist h_1, h_2 such that $h_1 \vDash_{\mathcal{V}} (W \triangleright [S])$, $h_2 \vDash (W \triangleright [T])$, and $h_1 \oplus h_2 \sqsubseteq h$. This means $dom(h_1) = dom(h_2) = W$, $cod(h_1) \supseteq S$, $cod(h_2) \supseteq T$. Since $h_1 \oplus h_2$ is defined, $dom(h_1) \cap dom(h_2) = cod(h_1) \cap cod(h_2)$, namely $W = cod(h_1) \cap cod(h_2) \supseteq S \cap T$. Now we take the conditionals c of $(id \otimes -\!\bullet_V)$, where V is the tensor product of objects in $cod(g)$ but not in W. It follow that f is equivalent to



So $(id_W \otimes -\!\bullet_V) \circ g$, h_1 and h_2 witness $W \perp\!\!\!\perp_{pr} S \mid T$. □

B Comparison with RV approach

This appendix is devoted to the connection between the string diagrammatic definitions of CI and the conditional independence for random variable from Simpson [20]. The goal is to provide a string diagrammatic notion of CI for random variables such that: (1) it covers all examples in Simpson [20], (2) random variables satisfying this notion also satisfies the axioms in Simpson [20], such as (local) independence structure and products.

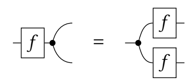
As for independence, some extra conditions always carries RV-CI structure. Moreover, this fragment is interesting and broad enough to cover the concrete examples in [20].

Before going into the details we recall some intuition of AS-CI. We know from basic probability theory that a random variable (RV) is a mapping from a sample space Ω (thought of as the set of all possible outcomes for some random event) to a measurable space $\mathcal{E}$ in which the likelihood of outcomes in Ω can be measured, thus forming a probability distribution. For plain independence, two RVs $X_i : \Omega \to \mathcal{E}_i$ $(i = 1, 2)$ are independent if their joint distribution can be calculated simply in terms of their respective distributions (as a product). For conditional independence, given another RV $Y : \Omega \to \mathcal{E}$, knowledge about the distribution over $\mathcal{E}$ provides a way to 'backward update' the knowledge about the sample space Ω by 'inverting' Y. This information provided by such inversion of Y may make the connection between X_1 and X_2, in which case we say X_1 and X_2 are conditional independent given Y. As an example, let Ω_{coin} be the set of all possible outcomes of flipping a fair coin three times, Y is the outcome of the second coin, X_1 is the outcome of the first and the second coins, X_2 is the outcome of the second and third coin. Then X_1 and X_2 are independent given Y.

Now we think of morphisms in an arbitrary category as abstract variables, and conditional independence of RVs now become a property on some set of abstract variables with the same domain. AS-CI in [20] is about so-called *multispans* in some $\mathbf{C}$ of the form $\{f_i : X \to Y_i\}_{i \in \mathbf{I}}$ where $\mathbf{I}$ is a finite index set, and each f_i is a $\mathbf{C}$-morphism. We will refer the readers to [20] for most concepts which are not explained here: multispans, (local) independence structure, (local) indenpendence product. Yet we will explain what does these concepts mean in the concrete setting. [Tao: make this precise]

Throughout this section we fix a Markov category $\mathbf{C}$. We will be interested in (subcategories of) the coslice category $A/\mathbf{C}$ for some object A, whose objects are $\mathbf{C}$-morphisms $A \to X$ and morphisms from $f : A \to X$ to $g : A \to Y$ are $\mathbf{C}$-morphisms $h : X \to Y$ such that $h \circ f = g$. Given $f_1 : X \to Y_1$ and $f_2 : X \to Y_2$, we write $f_1 \triangle f_2$ for the morphism $(f_1 \otimes f_2) \circ -\!\!\bullet\!\!<_X : X \to Y_1 \otimes Y_2$. This notation extends naturally to $\triangle_{i=1}^{k} f_i$ where each f_i is of type $X \to Y_i$, and it is well-defined by the associativity of $-\!\!\bullet\!\!<$.

Definition B.1. We call a $\mathbf{C}$-morphism f *deterministic* if it commutes with the copier:



Using the $\triangle$ notation, f is deterministic if $-\!\!\bullet\!\!< \circ f = f \triangle f$.

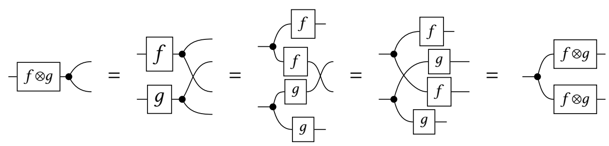
A multicategory $\mathcal{I}$ is a collection of multispans which has all singleton multispans and compositions of multispans that respect certain condition.

The category $\mathbf{E}$ of interest is the subcategory of $I/\mathbf{C}$ has all whose morphisms are deterministic. Indeed $\mathbf{E}$ still has enough structure for a string diagrammatic presentation.

Lemma B.2. *$\mathbf{E}$ is a Markov category.*

Lemma B.2. We need to show that $\mathbf{E}$ is a SMC, has copier and discarder compatible with the SM structure, and the discarder is terminal.

We define the symmetric monoidal structure on $\mathbf{E}$ to be $(\boxplus, J)$, where $(X, \omega_X) \boxplus (Y, \omega_Y)$ is $(X \otimes Y, \omega_X \otimes \omega_Y)$, and J is (I, id_I). That $(\boxplus, J)$ satisfies the symmetric monoidal structure conditions is inherited from that of $(\otimes, I)$. For example, $(X, \omega_X) \boxplus (I, id_I) = (X \otimes I, \omega_X \otimes id_I) \cong (X, \omega_X)$. So with a bit abuse of notations we use $\otimes$ for this monoidal product $\boxplus$ on $\mathbf{E}$ as well. Moreover, we need to verify the closure condition that if f, g are deterministic and admit Bayesian inversion, then $f \otimes g$ is also deterministic and admits Bayesian inversion (thus in $\mathbf{E}$). It is easy to see that $f \otimes g$ is also deterministic:

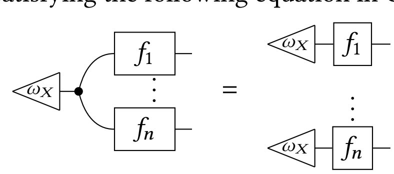


The CD structure is the same as that for $\mathbf{C}$: the copier $-\!\!\bullet\!\!<_X : (X, \omega_X) \to (X \otimes X, \omega_X \otimes \omega_X)$, and $-\!\bullet_X : (X, \omega_X) \to (I, id_I)$. The discarder is terminal essentially because $-\!\bullet$ is terminal in $\mathbf{C}$. □

We focus on two central criteria for (conditional) independence in [20]: (local) independence structure and (local) independent product. We will not repeat their full definitions from [20] here, and instead we will state our results and explain what the terminologies mean in our setup.

B.1 Independence

We define the set $\mathcal{I}$ of independent multispans on $\mathbf{E}$ to consist of $\{f_i : (X, \omega_X) \to (Y, \omega_Y)\}_{i \in \mathbf{I}}$ for some finite index set $\mathbf{I} = \{1, \ldots, n\}$ satisfying the following equation in $\mathbf{C}$:



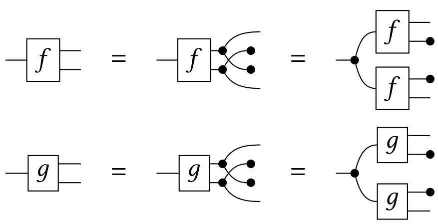
(23)

Lemma B.3. *The set of projections $\{\pi_i : (\bigotimes_{i \in \mathbf{I}} Y_i, \omega_{\otimes}) \to (Y_i, \omega_{Y_i})\}_{i \in \mathbf{I}}$ is jointly monic in $\mathbf{E}$.*

Proof. Given two $\mathbf{E}$-morphisms $f, g\colon (X, \omega_X) \to (\bigotimes Y_i, \omega_\otimes)$ satisfy that $\pi_i \circ f = \pi_i \circ g$ for all $i \in \mathbf{I}$, we show that $f = g$. We further assume that $\mathbf{I} = \{1, 2\}$, and that for arbitrary finite $\mathbf{I}$ is morally the same. $\pi_i \circ f = \pi_i \circ g$ for $i = 1, 2$ means


(24)

Then f, g being deterministic entails that they can be represented by their 'marginalisations':


(25)

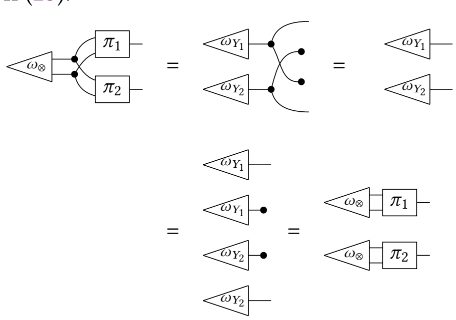
Applying (24) to (25), one immediately gets $f = g$. □

Lemma B.4. *If* $\{f_i\colon (X, \omega_X) \to (Y_i, \omega_{Y_i})\}_{i\in\mathbf{I}} \in \mathcal{I}$, *then* $\triangle_{i\in\mathbf{I}} f_i$ *is a* $\mathbf{E}$*-morphism* $(X, \omega_X) \to (\bigotimes_{i\in\mathbf{I}} Y_i, \bigotimes_{i\in\mathbf{I}} \omega_{Y_i})$.

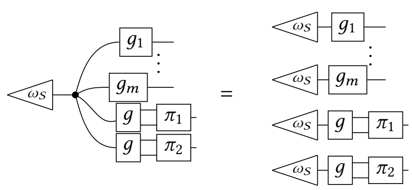
Proof. For readability we will omit the subscripts in the remainder of the proof. We first check that the type is correct, but $\triangle f_i \circ \omega_X = \bigotimes \omega_{Y_i}$ is exactly the independence condition. It follows immediately from all f_i's being deterministic that $\triangle f_i$ is also deterministic. □

Proof. B.6 For readability we write $\bigotimes Y_i$ for $\bigotimes_{i\in\mathbf{I}} Y_i$ in the remainder of the proof. According to Definition 3.5 in [20], we need to verify three facts: (1) $\mathbf{Proj} \coloneqq \{\pi_i\colon (\bigotimes_{i\in\mathbf{I}} Y_i, \omega_\otimes) \to (Y_i, \omega_{Y_i})\}_{i\in\mathbf{I}}$ is independent, (2) is $\mathcal{I}$-neutral, (3) if $\perp\!\!\!\perp \{f_i\colon (X, \omega_X) \to (Y_i, \omega_{Y_i})\}_{i\in\mathbf{I}}$ then there exists a unique $\mathbf{E}$-morphism $(f_i)_{i\in\mathbf{I}}\colon (X, \omega_X) \to (\bigotimes Y_i, \omega_\omega)$ such that $\pi_i \circ (f_i)_{i\in\mathbf{I}} = f_i$ for all $i \in \mathbf{I}$. In this proof we only consider $\mathbf{I} = \{1, 2\}$, and that for arbitrary finite $\mathbf{I}$ is morally the same.

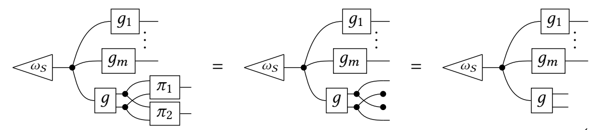
1. We show that **Proj** satisfies the independence condition (23).



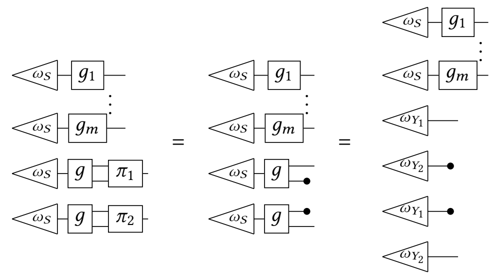
2. Suppose $\{g_j\colon (S, \omega_S) \to (T_j, \omega_{T_j})\}_{j\in\mathbf{J}} \cup \{g\colon (S, \omega_S) \to (\bigotimes Y_i, \omega_\otimes)\}$ satisfies that $\{g_j\}_{j\in\mathbf{J}} \cup \{\pi_i \circ g\}_{i=1,2}$ is in $\mathcal{I}$, where $\mathbf{J} = \{1, \ldots, m\}$. This means

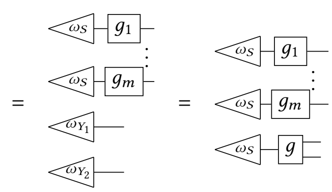

(26)

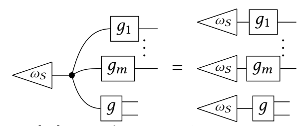
On one hand, since g is deterministic, the left-hand-side of (26) is equivalent to

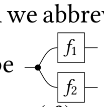

(27)

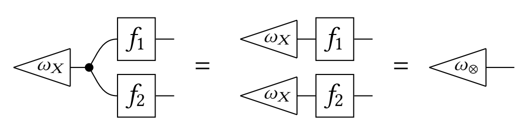
On the other hand, g satisfies that $g \circ \omega_S = \omega_\otimes = \omega_{Y_1} \otimes \omega_{Y_2}$, so the right-hand-side of (26) is


(28)



Then (26) implies , and by definition $\{g_j\}_{j\in\mathbf{J}} \cup \{g\}$ is also in $\mathcal{I}$.

3. Suppose $\{f_i\colon (X, \omega_X) \to (Y_i, \omega_{Y_i})\}_{i=1,2} \in \mathcal{I}$, we define $(f_i)_{i=1,2}\colon (X, \omega_X) \to (\bigotimes Y_i, \omega_\otimes)$ (which we abbreviate as (f_i) in the remainder of the proof) to be , and it is well-defined (namely (f_i) satisfies that $(f_i) \circ \omega_X = \omega_\otimes$) precisely because $\{f_1, f_2\} \in \mathcal{I}$:



It is clear that $\pi_i \circ (f_i) = f_i$, for $i = 1, 2$. For uniqueness, suppose $h\colon (X, \omega_X) \to (\bigotimes Y_i, \omega_\otimes)$ also satisfies $\pi_i \circ h = f_i$ for $i = 1, 2$. Then $\pi_i \circ h = \pi_i \circ (f_i)$ for $i = 1, 2$, and Lemma B.3 implies that $h = (f_i)$.

Therefore we conclude that $(\bigotimes_{i\in\mathbf{I}} Y_i, \omega_\otimes)$ together with $\{\pi_i\}_{i\in\mathbf{I}}$ form an independent product. □

Proposition B.5. $\mathcal{I}$ *has independence structure.*

Proposition B.5. It suffices to check that $\mathcal{I}$ is an affine multicategory of multispans satisfying that every singleton family $\{f\colon (X, \omega_X) \to (Y, \omega_Y)\}$ is in $\mathcal{I}$.

To see it is affine, given a multispan $\{f_i\colon (X, \omega_X) \to (Y_i, \omega_{Y_i})\}_{i\in\mathbf{I}}$ and an injective function $h\colon \mathbf{J} \to \mathbf{I}$, the set $\{f_{h(j)}\colon (X, \omega_X) \to (Y_{h(j)}, \omega_{Y_{h(j)}})\}_{j\in\mathbf{J}}$ satisfies (23) by discarding those $i \in \mathbf{I}$ not in the image of h. For singleton families $\{f: (X, \omega_X) \to (Y, \omega_Y)\}$, condition (23) holds trivially. □

This means that $\mathcal{I}$ has all singleton sets $\{f: (X, \omega_X) \to (Y, \omega_Y)\}$, for which (23) holds trivially. More interestingly, there exists so-called independent products satisfying some universal mapping property.

Proposition B.6. E *with independent structure* $\mathcal{I}$ *has* $\mathbf{I}$*-indexed independent products. In particular, given* $\{(Y_i, \omega_{Y_i})\}_{i\in\mathbf{I}}$ *with* $\mathbf{I} = \{1, \dots, n\}$*, the tensor product* $(\bigotimes_{i\in\mathbf{I}} Y_i, \omega_\otimes)$ *and the set of canonical projections* $\{\pi_i: (\bigotimes_{i\in\mathbf{I}} Y_i, \omega_\otimes)\}_{i\in\mathbf{I}}$ *are independent product and multispan of projections, where* $\omega_\otimes$ *is* $\omega_{Y_1} \otimes \cdots \otimes \omega_{Y_n}$.

Crucially, $(\bigotimes_{i\in\mathbf{I}} Y_i, \omega_\otimes)$ being an independent product with $\{\pi_i: (\bigotimes_{i\in\mathbf{I}} Y_i, \omega_\otimes)\}_{i\in\mathbf{I}}$ being a multispan of projections requires the following universal mapping property: if $\{f_i: X \to Y_i\}_{i\in\mathbf{I}} \in \mathcal{I}$, then there exists a *unique* **E**-morphism $f: (X, \omega_X) \to (\bigotimes Y_i, \omega_\otimes)$ such that $\pi_i \circ f = f_i$ for all $i \in \mathbf{I}$.

We can restate some examples in [20] in terms of some appropriate **E**.

Example B.7 (Finite probability distributions)**.** Example 2.1 from [20] is about independence in the category **FinProb**: the objects are pairs (X, p_X) where X is a finite set and $p_X: X \to (0, 1]$ satisfies $\sum_{x\in X} p_X(x) = 1$. Let **C** be $\mathcal{K}\ell(\mathcal{D})$, and **E** has as objects (X, ω_X) where $\omega_X: \mathbf{1} \to \mathcal{D}(X)$. The **E**-morphisms $f: (X, \omega_X) \to (Y, \omega_Y)$ satisfies that (1) $f(x)$ is of the form $|y\rangle$ (deterministic), (2) for all $y \in Y$ there exists finitely many $x \in X$ with $f(x) = |y\rangle$ (admits Bayesian inversions). **FinProb** isomorphic to a full subcategory of this **E** with restricted to those (X, ω_X) with $supp(\omega_X) = X$, denoted as **A**. Then the independence condition (23) of $\{f_i: (X, \omega_X) \to (Y_i, \omega_{Y_i})\}_{i=1,\dots,n}$ boils down to that, for all $(y_1, \dots, y_n) \in Y_1 \times \cdots \times Y_n$,

$$\sum_{x\in\bigcap\{f_i^{-1}(y_i)\,|\,i=1,\cdots,n\}} \omega_X(x) = \prod_{i=1,\dots n} \omega_{Y_i}(y_i)$$

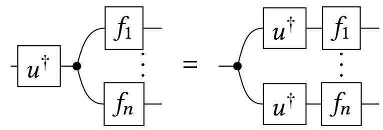
This is exactly the definition of independence of random variables in Example 2.1 from [20].

B.2 Conditional independence

As discussed at the beginning of Section B, for conditional independence between random variables one need some sort of backward update. Thus we require further that **C** admits *Bayesian inversion*:

[6, 10] The *Bayesian inversion* of a $I/\mathbf{C}$-morphism $f: (X, \omega_Y) \to (Y, \omega_Y)$ is a $I/\mathbf{C}$-morphism $f^\dagger: (Y, \omega_Y) \to (X, \omega_X)$ such that $f \circ f^\dagger = id_{(Y, \omega_Y)}$. We say f *admits Bayesian inversions* if such $f^\dagger$ exists.

Fix some **E**-object (U, ω_U), we define $\mathcal{I}_U$ be the family of multispans $\{f_i: (X, \omega_X) \to (Y_i, \omega_{Y_i})\}_{i\in\mathbf{I}}$ (for some finite index set $\mathbf{I} = \{1, \dots, n\}$) in the slice category $(U, \omega_U)\backslash\mathbf{E}$ (with a fixed $u: (X, \omega_X) \to (U, \omega_U)$) satisfying the following equation in **C**



(29)

where $u^\dagger: (U, \omega_U) \to (X, \omega_X)$ is the Bayesian inversion of $u: (X, \omega_X) \to (U, \omega_U)$. We will refer to the left-hand-side and the right-hand-side of (29) as s_1 and s_2, respectively.